

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 21:52:22 ; Search time 2461.97 Seconds  
(without alignments)  
13146.614 Million cell updates/sec

Title: PCT-US01-13059-3  
Perfect score: 3424  
Sequence: 1 aagcttaagctaccaacat.....aacggcagctagctgactcagt 3424

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_estl1:\*  
2: gb\_estl2:\*  
3: gb\_estl3:\*  
4: gb\_estl4:\*  
5: gb\_estl5:\*  
6: gb\_estl6:\*  
7: gb\_estl7:\*  
8: gb\_estl8:\*  
9: gb\_estl9:\*  
10: gb\_estl10:\*  
11: gb\_estl11:\*  
12: gb\_estl12:\*  
13: gb\_estl13:\*  
14: gb\_estl14:\*  
15: gb\_estl15:\*  
16: gb\_estl16:\*  
17: gb\_estl17:\*  
18: gb\_estl18:\*  
19: gb\_estl19:\*  
20: gb\_estl20:\*  
21: gb\_estl21:\*  
22: gb\_estl22:\*  
23: gb\_estl23:\*  
24: gb\_estl24:\*  
25: gb\_estl33:\*  
26: gb\_estl34:\*  
27: gb\_estl35:\*  
28: gb\_estl36:\*  
29: gb\_estl37:\*  
30: gb\_estl38:\*  
31: gb\_estl39:\*  
32: gb\_estl40:\*  
33: gb\_estl41:\*  
34: gb\_estl42:\*  
35: gb\_estl43:\*  
36: gb\_estl44:\*  
37: gb\_estl45:\*  
38: gb\_estl46:\*  
39: gb\_estl47:\*  
40: gb\_estl48:\*  
41: gb\_estl49:\*  
42: gb\_estl50:\*  
43: gb\_estl51:\*

em\_esthum10:\*  
44: em\_esthum11:\*  
45: em\_esthum12:\*  
46: em\_esthum13:\*  
47: em\_esthum14:\*  
48: em\_esthum15:\*  
49: em\_esthum16:\*  
50: em\_esthum17:\*  
51: em\_esthum18:\*  
52: em\_esthum19:\*  
53: em\_esthum20:\*  
54: em\_esthum21:\*  
55: em\_esthum22:\*  
56: em\_esthum23:\*  
57: em\_esthum24:\*  
58: em\_esthum25:\*  
59: em\_esthum26:\*  
60: em\_esthum27:\*  
61: em\_esthum28:\*  
62: em\_estin1:\*  
63: em\_estin2:\*  
64: em\_estin3:\*  
65: em\_estin4:\*  
66: em\_estin5:\*  
67: em\_estin6:\*  
68: em\_estin7:\*  
69: em\_estin8:\*  
70: em\_estin9:\*  
71: em\_estin10:\*  
72: em\_estin11:\*  
73: em\_estin12:\*  
74: em\_estin13:\*  
75: em\_estin14:\*  
76: em\_estin15:\*  
77: em\_estin16:\*  
78: em\_estin17:\*  
79: em\_estin18:\*  
80: em\_estin19:\*  
81: em\_estin20:\*  
82: em\_estin21:\*  
83: em\_estin22:\*  
84: em\_estin23:\*  
85: em\_estin24:\*  
86: em\_estin25:\*  
87: em\_estin26:\*  
88: em\_estin27:\*  
89: em\_estin28:\*  
90: em\_estin29:\*  
91: em\_estin30:\*  
92: em\_estin31:\*  
93: em\_estin32:\*  
94: em\_estin33:\*  
95: em\_estin34:\*  
96: em\_estin35:\*  
97: em\_estin36:\*  
98: em\_estin37:\*  
99: em\_estin38:\*  
100: em\_estin39:\*  
101: em\_estin40:\*  
102: em\_estin41:\*  
103: em\_estin42:\*  
104: em\_estin43:\*  
105: em\_estin44:\*  
106: em\_estin45:\*  
107: em\_estin46:\*  
108: em\_estin47:\*  
109: em\_estin48:\*  
110: em\_estin49:\*  
111: em\_estin50:\*  
112: em\_estin51:\*  
113: em\_estin52:\*  
114: em\_estin53:\*  
115: em\_estin54:\*  
116: em\_estin55:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	122.8	3.6	1337	81	BF630719	HVSMEB001
C 2	119.8	3.5	1280	80	BF264950	HV_CEA001
C 3	119.4	3.5	1101	219	CNS0021J	AL061936 Drosophil
C 4	117.2	3.4	1101	219	CNS00EVL	AL069706 Drosophil
C 5	114.4	3.3	1101	219	CNS0039G	AL063921 Drosophil
C 6	110.8	3.2	1219	80	BF264728	BF264728 HV_CEA001
C 7	104.8	3.1	829	221	CNS03LUB	AL249932 Tetraodon
C 8	101.6	3.0	1101	219	CNS00EVL	AL069706 Drosophil
C 9	101.2	3.0	1092	220	CNS0020K7	AL175696 Tetraodon
C 10	101.2	3.0	1101	219	CNS00E07	AL069440 Drosophil
C 11	100.6	2.9	1021	219	CNS014DY	AL104032 Drosophil
C 12	100.2	2.9	1101	219	CNS00LT7	AL078714 Drosophil
C 13	99.8	2.9	997	219	CNS005TE	AL060767 Drosophil
C 14	99.8	2.9	1067	80	BF264585	BF264585 HV_CEA000
C 15	99.6	2.9	1366	80	BF262464	BF262464 HV_CEA000
C 16	99.2	2.9	876	219	CNS000G1	AL053529 Drosophil
C 17	98.4	2.9	1032	220	CNS020IP	AL206746 Tetraodon
C 18	97.8	2.9	1135	220	CNS033G0	AL226115 Tetraodon
C 19	97	2.8	1092	220	CNS020K7	AL175696 Tetraodon
C 20	96.8	2.8	994	221	CNS04NOJ	AL298972 Tetraodon
C 21	96.4	2.8	1101	219	CNS0021J	AL061936 Drosophil
C 22	95	2.8	974	219	CNS00ITT	AL075432 Drosophil
C 23	94.8	2.8	1101	219	CNS017KE	AL108152 Drosophil
C 24	94.4	2.8	1201	219	CNS0167M	AL106396 Drosophil
C 25	94.2	2.8	935	256	B10881	B10881 P24H6-Sp6.1
C 26	94.2	2.8	1106	80	BF264948	BF264948 HV_CEA001
C 27	94	2.7	661	220	CNS02QVJ	AL209800 Tetraodon
C 28	94	2.7	945	221	CNS04DOK	AL285149 Tetraodon
C 29	94	2.7	1225	219	CNS0161D	AL106171 Drosophil
C 30	93.8	2.7	1101	219	CNS00E07	AL069440 Drosophil
C 31	93.6	2.7	1101	219	CNS00EYC	AL069603 Drosophil
C 32	93.4	2.7	576	220	CNS035N7	AL228940 Tetraodon
C 33	93.4	2.7	848	245	AZ549808	AZ549808 ENTF262TF
C 34	93.4	2.7	1190	220	CNS020N7	BF264865 HV_CEA001
C 35	93.2	2.7	1007	80	BF264865	BF264865 HV_CEA001
C 36	93.2	2.7	1073	146	BF274651	BF274651 GA_EB002
C 37	93.2	2.7	1248	256	B11336	B11336 F19M10-Sp6
C 38	92.8	2.7	868	219	CNS010YV	AL099601 Drosophil
C 39	92.6	2.7	625	220	CNS036A2	AL229763 Tetraodon
C 40	92.2	2.7	524	220	CNS01U90	AL167541 Tetraodon
C 41	92.2	2.7	1101	219	CNS003BD	AL064091 Drosophil
C 42	92.2	2.7	1101	219	CNS003DQ	AL064580 Drosophil
C 43	92	2.7	865	146	BF264686	BF264686 HV_CEA001
C 44	92	2.7	1101	219	CNS006TE	AL065901 Drosophil
C 45	91.8	2.7	1029	220	CNS01ZGM	AL174271 Tetraodon

ALIGNMENTS

RESULT 1	
BF630719/c	
ID BF630719	standard; RNA; EST; 1337 BP.
XX	
AC - BF630719;	
XX	
SV BF630719.1	
XX	
DT 21-DEC-2000 (Rel. 66, Created)	
DT 21-DEC-2000 (Rel. 66, Last updated, Version 1)	
XX	
DE HVSMEB0013H16f Hordeum vulgare seedling shoot EST library HVCNDA0002	
DE (dehydration stress) Hordeum vulgare cDNA clone HVSMEB0013H16f, mRNA	
DE sequence.	
XX	
KW EST.	
XX	

OS	Hordeum vulgare (barley)	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; Liliopsida; Poales; Poaceae; Triticeae; Hordeum.	
XX		
RN	[1]	
RP	1-1337	
RA	Wing R., Close T.J., Kleinbartsch A., Wise R., Begum D., Frisch D., Yu Y.,	
RA	Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,	
RA	Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;	
RA	"Development of a genetically and physically anchored EST resource for	
RA	barley genomics";	
RL	Unpublished.	
XX		
CC	Contact: Wing RA	
CC	Clemson University Genomics Institute	
CC	Clemson University	
CC	100 Jordan Hall, Clemson, SC 29634, USA	
CC	Tel: 864 656 7288	
CC	Fax: 864 656 4293	
CC	Email: rwing@clemson.edu	
CC	Seq primer: AATTACCCCTCACTAAAGG	
CC	High quality sequence start: 38	
CC	High quality sequence stop: 1204.	
XX		
XX	Key	Location/Qualifiers
FT	source	1..1337
FT	/db_xref="taxon:4513"	
FT	/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"	
FT	/organism="Hordeum vulgare"	
FT	/cultivar="Morex"	
FT	/clone="HVSMEB0013H16f"	
FT	/clone_lib="Hordeum vulgare seedling shoot EST library	
FT	HVCNDA0002 (Dehydration stress)"	
FT	/tissue_type="Seedling shoot"	
FT	/lab_host="TJC121"	
XX		
SQ	Sequence 1337 BP; 292 A; 22 C; 74 G; 941 T; 8 other;	
	Query Match	3.6%; Score 122.8; DB 81; Length 1337;
	Best Local Similarity	45.7%; Pred. No. 8.3e-11;
	Matches 535; Conservative	0; Mismatches 623; Indels 12; Gaps 3;
QY	557 aaagaagaagaatcatattactattatagaattaaagaagcagtgtaaaacgtgaa	616
Db	1197 ATAAATNAATNAATNAATNAATNAATNAATNAATNAATNAATNAATNAAT	1138
QY	617 agcagaattttctaaaaaaaataagtaaacgtctacaaacttatttattgtgtataacata	676
Db	1137 TAAATNAATNAATNAATNAATNAATNAATNAATNAATNAATNAATNAAT	1078
QY	677 tctataagaagaactcaatatatgataaatcattttcaacaaaatttctatgaattataa	736
Db	1077 TAGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1021
QY	737 taaaaaagtcattttgacacttaaaaggttgacataaacccgtctctccaaaaaaat	796
Db	1020 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	961
QY	797 caaaacattataattttctaaaaactatggtgtaattttgtgaaatcaaaagaaagaa	856
Db	960 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	901
QY	857 ggatttctatcatcaagtttctatttattgtatcaaaactttcaaaatttctgaattgaa	916
Db	900 AAAAGCAAAACAAAACAAATGATAGATTTAATACATAAATAAATAAATAAATAA	841
QY	917 aggaaaaaataagaataataatgtttgttttttttttttttttttttttttttttttt	976
Db	840 ATAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	781
QY	977 atattcatacaaaaaaatgtatttttaattatgatgatgagattaccatcccaaaagggtcgaac	1036

```
Db 780 AAATATATATAATA-----TAAATTAATAATACAAATAATTAATAATAATAATAATAATAATA 727
Qy 1037 ttataataaacaagttaactaaacatacatgtgatcaaatcaatgacagatttga 1096
Db 726 AAAAAATTAATAATAATACATAATAATAATAATAATAATAATAATAATAATAATAATAATA 667
Qy 1097 tcttaataatagaatgattgagcaaacctcaaaaatgtctcttaggacacaaatctt 1156
Db 666 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 607
Qy 1157 tcttttagcttattaaagccggaggttcaactctctccctgttagacttttggtttc 1216
Db 606 AAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 547
Qy 1217 aaatcttttttccaaaaaataatagtttaattgtagggcataatatttggtttaatt 1276
Db 546 AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 487
Qy 1277 aagtcctatagatttttaggaccatctctaatcacgacaaatatctccttaattgtaacaca 1336
Db 486 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 427
Qy 1337 tttaaaacttaaaagtattgcaattcacaatccttcaaatatataatatatatata 1396
Db 426 TAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 367
Qy 1397 t---atatatatatatatagaagtatatagaacgataactccttactcaacaat 1453
Db 366 TAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 307
Qy 1454 tagcccaaaaaaacatccatgatttaactaggaatttttaacaaactcaaataggt 1513
Db 306 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 247
Qy 1514 tggtagttaaaaaaacaataatagtagtgatcacatgacaccttttaaaaaatatata 1573
Db 246 TATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 187
Qy 1574 tatcgaagaagtttttaatttggcaattataatatacatctctatcaatttaaatatatt 1633
Db 186 RACCTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 127
Qy 1634 aataatgcataattctgtaatatctatcttcttaattccatagacacacaaataaa 1693
Db 126 ATCAGTAGCAGTACTCANTATCTAATAACAGAAAAATAGAAAAATAATAATAATAATA 67
Qy 1694 acatatcaaatagtttttaacttaacaaataa 1723
Db 66 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 37

RESULT 2
BF264950
ID BF264950 standard; RNA; EST: 1280 BP.
XX AC BF264950;
XX SV BF264950.1
XX DT 20-NOV-2000 (Rel. 65, Created)
XX DT 20-NOV-2000 (Rel. 65, Last updated, Version 1)
XX DE HV_CEA0010N09f Hordeum vulgare seedling green leaf EST library HVCDNA0004
XX DE (Erysiphe infected & control) Hordeum vulgare cDNA clone HV_CEA0010N09f,
XX DE mRNA sequence.
XX KW EST.
XX OS Hordeum vulgare (barley)
XX OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
XX OC Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
```

```

[1]
RP 1-1280
RA Wing R., Close T.J., Kleinhofs A., Wise R., Bequm D., Frisch D., Yu Y.,
RA Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,
RA Sasaki C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;
RA "Development of a genetically and physically anchored EST resource for
RA barley genomics";
RL Unpublished.
XX
CC Contact: Wing RA
CC Clemson University Genomics Institute
CC Clemson University
CC 100 Jordan Hall, Clemson, SC 29634, USA
CC Tel: 864 656 7288
CC Fax: 864 656 4293
CC Email: rwing@clemson.edu
CC Seq primer: AATTAAACCTCCTCAATAAGGG
CC High quality sequence start: 39
CC High quality sequence stop: 1197.
XX
XX Key Location/Qualifiers
XX source 1. 1280
XX /db_xref="taxon:4513"
XX /db_xref="ESTLIB:6818"
XX /note="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
XX /organism="Hordeum vulgare"
XX /cultivar="C116155 (M1313)"
XX /clone="HV_CEA0010N09f"
XX /clone_lib="Hordeum vulgare seedling green leaf EST library
XX HVCDNA0004 (Erysiphe infected & control)"
XX /tissue_type="seedling green leaf"
XX /lab_host="TJC121"
XX
SQ Sequence 1280 BP; 731 A; 15 C; 18 G; 353 T; 163 other;

Query Match 3.5%; Score 119.8; DB 80; Length 1280;
Best Local Similarity 43.9%; Pred. No. 2.6e-10;
Matches 471; Conservative 0; Mismatches 594; Indels 7; Gaps 4;

Qy 519 agaacagaccacagcaaatatttattgtattttctataaagaagaagaattcatatt 578
Db 212 AANNANANANNNNNNNNANNNANNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 271
Qy 579 actatagaattaaagctaaagctgaaacgtgaaagcagaattctctaaaaaata 638
Db 272 NAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAATAAAAAATAAAT 331
Qy 639 gtaactgtcacaaacttattgtgtatatacatatctataaagaagaactcaaatata 698
Db 332 AATTAATAATTTAAAAATAAAAAATATAAAAAAATATATATATATAATTAATANNANNA 391
Qy 699 tgataaatcatcttaacaaattctctatgaattataataaaaaaagtcacatttgacac 758
Db 392 AAAAAATTTAATTAATTTANNNTTAAAAATANNNAATTTAAANAAAAAATAAAAAAN 451
Qy 759 ttaaaaggttgacaataaccgtctctccaaaaaatacaaacatttataattctcaaa 818
Db 452 TTAANAAAAAANAAAAAANAAAAAATAAAAAANAAAAAANAAAAAANAAAAAANAAAA 511
Qy 819 actatggtgtaatttctgctgaatcaaaagaagaaggatttcttatcatcatttc 878
Db 512 AAAAAAANAAAAAANAAAAAANAAAAAATTAATTAATAAAAAAANAAAAAATATAA 571
Qy 879 attattgtcaaaacttcaaatcttgatgttaattgaaaggaaagaaataaagataat 938
Db 572 ATAAAAAATTTAAAAATAAAAAAANAAAAAATAAAAAATAAAAAATANNNAATTTATT 631
Qy 939 gttgtttttgttcttattgttacattttcttggaattatattcatcaacaaatgtgat 998
Db 632 TTTTATTATTTATATATATATATATATATATATATATATATATATATATATATATAT 687
```

[illegible]





[illegible]

RESULT	7
CNS03LUB	





Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE  
AUTHORS  
1 (bases 1 to 1092)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billaut,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 1092)  
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
3 (bases 1 to 1092)  
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Unpublished  
Genoscope.  
3 (bases 1 to 1092)  
Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
source  
1. 1092  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="222L1"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG222CF06LP1-end : 77"

BASE COUNT  
383 a 169 c 165 g 262 t 113 others

ORIGIN  
Query Match 3.0%; Score 101.4; DB 220; Length 1092;  
Best Local Similarity 41.6%; Pred. No. 2.7e-07;  
Matches 207; Conservative 66; Mismatches 220; Indels 4; Gaps 1;  
QY 566 aagaattcattactatagaattaaagctgaagcagctgaaacgtgaaagcagaatt 625  
DB 600 MHAHHHTTTTAAAWTTATTTAAWWAAAAAATAAAATTTTAAAAA 659  
QY 626 tctaaaaaataagctgaacacttatttctgtatataacatctctataaag 685  
DB 660 AAATAAATTTTAAATAAAATTTTATTTTATTTTATTTTATTTTAA 719  
QY 686 aaactcaaatatgatgaatacattttcaacaaatttctatgaattataataaaag 745  
DB 720 AAATTAATTAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 779  
QY 746 tcactttgacactaaaggttgacaataaccgtctctcccaaaaaatacaaacatt 805  
DB 780 WTWTAAATTTAAWWATTAATTTATTTTAAATAAAATAAAATAAAATA 839  
QY 806 tataattctcaaacacttggtgtatttctgctgaatacaaaagaaagaggtttcta 865  
DB 840 AATAATTTTAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 895  
QY 866 tatcataaggttcattatgtatcaaaccttcaaatctcgaatttgaaagaaagaa 925  
DB 896 AATTTTATTTTAAATAAAATAAAATAAAATAAAATAAAATAAAATA 955  
QY 926 attagataatagttgtttgtttgtttgtttgtttgtttgtttgtttgttt 985  
DB 956 AATAATAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1015  
QY 986 acaaaaaatgtattttaaataatgatgagagattaccatccaaaggtcgaaattataaa 1045  
DB 1016 AATAAAATATTATATATATAATAATAATAATAATAATAATAATA 1075  
QY 1046 acaagtttaataactaaa 1062

Db 1076 AAAAAAATAAAATAAA 1092

RESULT 10  
CNS00E07  
LOCUS  
DEFINITION  
CNS00E07 1101 bp DNA GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION  
AL069440  
VERSION  
AL069440.1 GI:4949583  
KEYWORDS  
GSS.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
Genoscope.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACS. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mamoser in Piter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
PI and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source  
1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"  
/clone="BACR29P01"  
/note="end : TET3"

BASE COUNT  
366 a 66 c 104 g 351 t 214 others

ORIGIN  
Query Match 3.0%; Score 101.2; DB 219; Length 1101;  
Best Local Similarity 36.8%; Pred. No. 2.9e-07;  
Matches 230; Conservative 90; Mismatches 305; Indels 0; Gaps 0;  
QY 509 aagaaaaatcagaacagaccacgaattttatgtattttctattataaaaaagaagaag 568  
DB 417 AAAAAAATCAATAAAARRGGGCAATAATTTTGAACATATATAAAAAAATAAAT 476  
QY 569 aatcattacttactatagaattaaaacgtgaaacgtgaaacgtgaaacgtatctct 628  
DB 477 AATATTAAATTAATAATAAAATTTTAAAAAATAAAATAAAATAAAATAAA 536  
QY 629 aaaaaaatagtaactgcacaaactatttattgtatataacatctctataaagaa 688  
DB 537 TAATTTTATATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 596  
QY 689 ctcaaatatgatataatcttttaacaaatttctatgaattataataaaaaagttca 748  
DB 597 TATATATTAAAGWAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATA 656  
QY 749 cttttgacacttaaaaggttgacaataacgtctctcccaaaaaatacaaacattat 808  
DB 657 ATTTATTAAATTTAAATTTTAAATAATTAATAATTAATAATTAATAATTAATA 716

[illegible]

RESULT	11
CNS014DY/C	
LOCUS	CNS014DY      1021 bp      DNA      GSS      26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC BACN1004 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

**ACCESSION** AL104032  
**VERSION** AL104032.1 GI:5615643  
**KEYWORDS** GSS.

REINWORDS  
SOURCE  
ORGANISM  
Plasmid *Drosophila melanogaster*  
fruit fly.  
GSS.

ORGANISM  
Plasmodium falciparum  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE  
1 (bases 1 to 1021)  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE	AUTHORS	TITLE
1 (PAGES 1 TO 1021)	Genoscope.	Direct Submission

JOURNAL TITLE  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage  
Direct Submission  
RD 191 91006 Evry cedex - FRANCE (E-mail : seqtraf@genoscope.cns.fr)

BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

COMMENT  
Determination of this BAC-end sequence was carried out as part of collaboration with the European Drosophila Genome Project (EDGP)

<http://www.edgp.ebi.ac.uk> - This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre

anuary (Dios Brac) was made by Alain Billaud at CERH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Buchet.

project grant. The DNA was prepared from embryos by Alain Buchet and Genevieve Payan. It has been constructed in the vector pBelorac11.

```

FEATURES
source      1 1021
Location/Qualifiers
pbe10BAC11.

```

```
source
1. .1021
/organism="Drosophila melanogaster"
```

```
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"
```

```

/clone_lib="DrosBAC"
/clone="BACN11004"

```

```

/clone= BACN11004
/note="end : T7"
342 a 1 c 104 a 462 + 112 others
BASE COUNT

```

ORIGIN	BASE COUNT	I	C	G	A	T	others
	342	104	462	112			

[illegible]

Query Match 2.9%; Score 100.6; DB 219; Length 1021;  
Best Local Similarity 40.8%; Pred. No. 3.7e-07;

Matches 414; Conservative 78; Mismatches 512; Indels 10; Gaps
---

Qy 725 atgaattataataaaaaagtcacitttgacacttaaaagttgacaataaacgtctct 784

Db 1018 AASAAACAAAAAHHAAAAHAWAAAAAWTTTTTTTTTAAACMAAAAWTTTTTTT 959

Qy	785	ccaaaaaaatcaaacatttataattcttcaaaactatggtgtaatttggctgaaatca	844
Db	958	WMTTTTTTTTTTTWAAAYWMTCCWTTTTCMCACAAAATWTATAWTAATWTWACCMAAA	899
Qy	845	aaagaaaaagaagattctctatcatcataagttctcaattgtatcaaacatttcaaatctc	904
Db	898	ATHAMAAAACACCMACMAAAATWTTAAWATWMAAAATAAAATAAAAAAATWMTWMTAW	839
Qy	905	atgtaattggaaggaaaaaatcaagatat--aatgttgtttgttgttcttctgtttaca	962
Db	838	ATWTCATCCCAATTAANTCAAWTAACWCCCAAAAAATAAAWAAATWHAYTAACAAAA	779
Qy	963	ttttcgtgaatatattctatacaaaaaatgatatattttaatatgatgagagataccat	1022
Db	778	TAAAYCATACMAAAYAAACCAATWMTATATTTAACACCMAMTWMACCAACAACACAC	719
Qy	1023	ccaaaagtcgaacttatataaacaagttataataactaaacaatacatgatgcacaact	1082
Db	718	ACWATMCACATCTCTTATCCACACACTGMAAAACANTTACTCANWAMATATWCTTACTT	659
Qy	1083	aatgacagtttgtatcttaaaatagaaatgattgagcaaac--tcaaaaaatgtctcttta	1141
Db	658	ACAAVAMTCCACMWTAAAAACAAAMWATATCCTTCACCCTTWWAAAAWAAAWATATA	599
Qy	1142	ggatcacaataatcttctcttgagctattataagcgggaggttcaactctctctccctgt	1201
Db	598	TCMYCACCTTTTCAACAAATWAAWTATTATTACCAHTTTTAWTTAWATRTWMTAACCTTC	539
Qy	1202	agacttttgtttca--aatcttttcttcaaaaaatcaataattagttaattggcca	1258
Db	538	TCCAYTAAACATMTACTTAYTHCTTWCCTCCCAAAAMTMTYCCAAATTCATPCCWATC	479
Qy	1259	taattttgttttaataagtcocatagatttttttagaccatctcttaatcacgacaaat	1318
Db	478	HTTTTAATWCTTCCAMTCCACCCCAAAWAAATTAATAAAATAATATTATTATTTAAAA	419
Qy	1319	atcct----aaatgtacaacatttaaaacttaaaagattgtcattcaacaatcttcaaaa	1374
Db	418	TTWATTTTAAAAATATTAAATTTTAAAAATATAATAATTTTTTTTTTAAAAAATTTAAA	359
Qy	1375	tatatatatatatatatatatatatatatatatataatgaaagtttatagaaacga	1434
Db	358	TTTATTTATTTTTTAAAAATTTTTCAAATTAATTTTTTTTTTTTTTAAATTTTTTATTNATC	299
Qy	1435	taactcttactcaacaatttagcccaaaaaacatcccataatgcatttaaactagggaatt	1494
Db	298	TAWATATTAAATAAATAATATTTTTTTTATAATTAATAATAAAWTTTTTTAAAAAATAAT	239
Qy	1495	ttaacaactcaaataggttggtagttaaaaaaaacaaatagtagatgcatacatcgtac	1554
Db	238	TAAAAATTTTTTATATATTTTTTAAATTTTTTAAATTAATAAAAAAATAATAAAATATT	179
Qy	1555	cttttaaaatatatactcatcgaaagttttcaaatgttcgaaatcaaatcacattctac	1614
Db	178	TTTTAACAAATATATTTTTTATTAATAAAATATTAACAAATTTTTTAAAAAATAATAATAA	119
Qy	1615	tatcaattaaaaacatttaataatgcaataatctgttaatatcttatcttcttaattccata	1674
Db	118	TWAAATTTATTTAAAWACAATTAATAWAAATTTTAAATTAANTTAATTTATATTAAAAA	59
Qy	1675	tagaaccaaaacaaaataacatatcaaatagtttttaactcaacaaaacogtta	1728
Db	58	AAAAAAAATATAAAAAATTTAAATATATTTAAATAATTAATAAATTAATATATAATTTTTT	5

[illegible]

AL078714.1	GI:5102004
GSS.	
fruit fly.	Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
1 (bases 1 to 1101)	
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (11-JUN-1999), Genoscope - Centre National de Sequencage ;
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with The Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila Melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html This library was prepared by Kazutoyo Osoegawa and Aaron Mammossin at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPcI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	Location/Qualifiers
source	i . 1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /cclone_lib="RPcI-98" /cclone="BACR48P19" /note="end c Ter3"
BASE COUNT	469 a     6 c     69 g     151 t     406 others
ORIGIN	
Query Match	2.9%; Score 100.2; DB 219; Length 1101;
Best Local Similarity	41.6% Pred. No. 4 3e-07;
Matches 229:	Conservative 66; Mismatches 252; Indels 4; Gaps 1;
QY	509 agaaaaatcagaaaggcacgcaaatatttatgttattttctttaaaaagaagaagaag 568 
Db	196 AAA 255
QY	569 aattcatattacttagaattaaaagtgaaccttgtaaacgtgaagcagaatttc 628 
Db	256 AAA 315
QY	629 aaaaaaatagtaaactgcacaaacttatttgtatatataacatctcataaaga 688 
Db	316 AAA 375
QY	689 ctcaaatatgatataatcattttaacaatttctatgaattataataaaaaaagtca 748 
Db	376 AAA 435
QY	749 cttttgaccttaaaggttigacaactcgctccataaaaaaaatcaaaccatttat 808 
Db	436 AAA 495
QY	809 aattctaaaactatggdgttaattttgctgaatacaaaagaagaagaagattctatat 868 :::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
Db	496 TTWKATTTTDTDAADDKWAAWDDTTTTTAATTWTWTTATAATTTTTTKKTTTW 555
QY	869 cataagttctcatattgtatcaaacattccaatttcattgaattgaaagaaagaatt 928    ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
Db	556 KAYKTDWKKAKTAANKTAANAWWTTTTTTTAAVVTWTTTKTKTWTAKKWAKWWDATTT 615
QY	929 aagatataatgttgt-----ttttgtttcttatattttcatgggaatatattcat 984 : : :   : :   : :   : :   : :   : :   : :   : :   : :   : :   :



```

QY 1151 aatcttctttagcttattaaagccggaggtagtcaactctctctccctgtagacttttt 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 413
QY - 1211 gtttcaaatcttcttcttcaaaaaaacaataatagtagtgagcgcaataatttggt 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 353
QY 1271 ttaataagtcacatagatttttttagaccat 1301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 AGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 322

RESULT 15
ID BF262464/c standard; RNA; EST; 1366 BP.
XX BF262464
AC BF262464;
XX BF262464.1
SV
XX
XX 20-NOV-2000 (Rel. 65, Created)
DT 20-NOV-2000 (Rel. 65, Last updated, Version 1)
XX
XX HV_CEA0004Allif Hordeum vulgare seedling green leaf EST library HVCDNA0004
DE (Erysiphe infected & control) Hordeum vulgare cDNA clone HV_CEA0004Allif,
DE mRNA sequence.
XX
XX KW EST.
XX
OS Hordeum vulgare (barley)
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Triticeae; Hordeum.
XX
XX [1]
RP 1-1366
RA Wing R., Close T.J., Kleinhofs A., Wise R., Begum D., Frisch D., Yu Y.,
RA Anderson H., Dale J., Henry D., Kernodle S., Palmer M., Rambo T.,
RA Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.;
RT "Development of a genetically and physically anchored EST resource for
RT barley genomics";
RL Unpublished.
XX
XX Contact: Wing RA
CC Clemson University Genomics Institute
CC Clemson University
CC 100 Jordan Hall, Clemson, SC 29634, USA
CC Tel: 864 656 7288
CC Fax: 864 656 4293
CC Email: rwing@clemson.edu
CC Seq primer: AATTAACCTCACTAAAGGG
CC High quality sequence start: 428
CC High quality sequence stop: 598.
XX
XX Key Location/Qualifiers
FH
FH source
FT 1. .1366
FT /db_xref="taxon:4513"
FT /db_xref="ESTLIB:6818"
FT /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
FT /organism="Hordeum vulgare"
FT /cultivar="C116155 (M1a13)"
FT /clone="HV_CEA0004Allif"
FT /clone_lib="Hordeum vulgare seedling green leaf EST library
FT HVCNA0004 (Erysiphe infected & control)"
FT /tissue_type="seedling green leaf"
FT /lab_host="TJC121"
XX
SQ Sequence 1366 BP; 260 A; 37 C; 0 G; 642 T; 427 other;

```

Query Match

2.9%; Score 99.6; DB 80; Length 1366;

Best Local Similarity 45.0%; Pred. No. 5.3e-07;  
Matches 407; Conservative 0; Mismatches 491; Indels 6; Gaps 2;

```

QY 672 acatactcataaagaacactcaaatatataatgataaatacatttttaacaaaattttctatgaat 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1323 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1264
QY 732 tataataaaaaagccacttttgacacttaaaaggtgacataa-----cgcctctccc 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1263 AAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1204
QY 787 aaaaaaaatcaaacattttataattttcaaacatggtgtaattttgctggaatcaaa 846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1203 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1144
QY 847 aaaaaagaagatttctatcataaagtttcatttattgtatcacaactttcacaatttccat 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1143 AATTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 1084
QY 907 gtaattgaaaggaaaaaataagataataatgtgtttttgtttcttctatgttccatttt 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1083 ATAAATTTATCGGTTATATATATATATATATATATATATATATATATATATATATAT 1024
QY 967 catggaatataatcaacaaataatgtattttaatatgatgagagattaccatccaa 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1023 ATAAAAATATATATATATATATATATATATATATATATATATATATATATATATAT 964
QY 1027 aagtcacaacttatataaaa--caagttaataactaaacaacacacatggtgacacacaa 1085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 963 TATTTAGATATATATATATATATATATATATATATATATATATATATATATATATAT 904
QY 1086 gacagttttgacttaaaatagaatgattgagcaaacctcaaaaatgtcttcttaggat 1145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 903 AATTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 844
QY 1146 cacaataatcttctttagcttataaagccggaggtagtcaactctctctccctgtagac 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 843 AAATAAAGATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 784
QY 1206 tttgttttcaaatcttcttcttcaaaaaatcaataatagtagtaattggcgcaataatt 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 783 TTTTAAATGATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 724
QY 1266 tggttttaattagtcacatagattttttaggaccatctcttaacacgacacaaatccctaa 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 723 NNATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 664
QY 1326 attgtaacacatttaaaacttaaaagattgtgattccacaatcccttaaaatatatatat 1385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 663 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 604
QY 1386 atatatatatatatatatatatatatgaaagtttatagaaacgataaactcccttac 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 544
QY 1446 tcaacaattagcccaaaaaaacatcccatgaatgctttaaacctgaagatttttaacaaactc 1505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 484
QY 1506 aaatagttggttagttaaaaaaaacaaatagtagatgtagatcgtagctaccttttaaaata 1565
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 AAAAAATATATATATATATATATATATATATATATATATATATATATATATATATAT 424
QY 1566 tata 1569
Db 423 ANNA 420

```

Search completed: July 12, 2001, 23:04:16  
Job time: 4314 sec

---

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 22:06:12 ; Search time 92.84 Seconds  
(without alignments)  
6832.523 Million cell updates/sec

Title: PCT-US01-13059-3  
Perfect score: 3424  
Sequence: 1 aagcttaagctaccacat.....aacggcagctgactcagct 3424

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	115.6	3.4	19124	2	US-08-487-826B-13
C 2	103.8	3.0	7218	1	US-08-232-463-14
C 3	83.4	2.4	19124	2	US-08-487-826B-13
C 4	79.4	2.3	8920	2	US-08-446-855A-1
C 5	79.4	2.3	8920	4	US-09-150-741-1
C 6	71.8	2.1	6152	4	US-08-973-462-1
C 7	71	2.1	4673	1	US-07-638-431-1
C 8	70.6	2.1	4673	5	PCT-US92-00018-1
C 9	69.4	2.0	51952	4	US-07-867-106-2
C 10	68.4	2.0	637	4	US-08-947-823-1
C 11	68.4	2.0	835	4	US-08-998-416-288
C 12	67.6	2.0	617	4	US-08-998-416-186
C 13	66.2	1.9	3095	6	5231168-1
C 14	65.2	1.9	665	2	US-08-883-795A-36
C 15	64.8	1.9	665	2	US-08-883-795A-36
C 16	64.6	1.9	6152	4	US-08-973-462-1
C 17	64.2	1.9	6768	1	US-08-107-755A-1
C 18	64.2	1.9	8457	1	US-07-991-867B-1
C 19	64.2	1.9	8457	2	US-08-544-332-1
C 20	64.2	1.9	8920	2	US-08-446-855A-1
C 21	64.2	1.9	8920	4	US-09-150-741-1
C 22	63.8	1.9	636	4	US-08-998-416-1137
C 23	63.8	1.9	701	4	US-08-998-416-701
C 24	63.4	1.9	834	4	US-08-998-416-305
C 25	63	1.8	51952	4	US-08-947-823-1
C 26	62.6	1.8	6243	2	US-09-056-075-1
C 27	62.4	1.8	724	4	US-08-998-416-693

28	62.4	1.8	732	4	US-08-998-416-1036	Sequence 1036, Ap
29	62.4	1.8	837	4	US-08-998-416-288	Sequence 288, App
C 30	61.6	1.8	731	1	US-08-451-405A-2	Sequence 2, Appli
31	60.8	1.8	636	4	US-08-998-416-1137	Sequence 1137, Ap
32	60.6	1.8	1511	1	US-07-991-867B-8	Sequence 8, Appli
33	60.6	1.8	1511	1	US-08-107-755A-8	Sequence 8, Appli
34	60.6	1.8	1511	2	US-08-544-332-8	Sequence 8, Appli
C 35	59.8	1.7	1850	4	US-08-617-860B-32	Sequence 32, Appli
C 36	59.8	1.7	4098	2	US-08-605-106-4	Sequence 4, Appli
C 37	59.6	1.7	3095	6	5231168-1	Patent No. 5231168
C 38	59	1.7	658	4	US-08-998-416-595	Sequence 595, App
39	58.8	1.7	1368	3	US-08-874-563-5	Sequence 5, Appli
40	58.8	1.7	1368	4	US-08-577-483-14	Sequence 14, Appl
C 41	58.8	1.7	3701	4	US-08-845-258-10	Sequence 10, Appl
C 42	58.8	1.7	3701	4	US-08-990-571-10	Sequence 10, Appl
C 43	58.8	1.7	4253	4	US-08-577-483-7	Sequence 7, Appli
44	58.6	1.7	711	4	US-08-998-416-786	Sequence 786, App
C 45	58.6	1.7	782	4	US-08-998-416-224	Sequence 224, App

ALIGNMENTS

RESULT 1  
US-08-487-826B-13/C  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israelson, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH121.001CPI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 3.4%; Score 115.6; DB 2; Length 19124;



```
RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

Query Match 3.0%; Score 103.8; DB 1; Length 7218;
Best Local Similarity 2.5%; Pred. No. 1.5e-10;
Matches 9; Conservative 257; Mismatches 99; Indels 0; Gaps 0;

QY 22 gaatttagtaaaagaccatgatttgaaattgtcgcaaatcgagaagatatata 81
||||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1445 GAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1386

QY 82 gagccgacaggggaacagtgaaacacacaaagcgcgtaagaatgaacagtgaggagaag 141
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1385 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1326

QY 142 aagagagaatctaccatcattcgagggaagaagatgggaatcagagaaaaatctgga 201
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1325 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1266

QY 202 aaagaaataaagagaaagagagaaagtgaggagaaagatgcagtgaaagactgct 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1265 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1206

QY 262 atagccacatccacatggtgtgtgatgagagagagagagagtgtaaaagcgaatt 321
```

```
Db 1205 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1146
QY 322 gtggagagataaagagagagagactgagcgagtgcaagtcgtcgtgtttaaaga 381
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1145 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1086
QY 382 aagaa 386
: : : : :
Db 1085 RRRRR 1081

RESULT 3
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnols, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92860
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match 2.4%; Score 83.4; DB 2; Length 19124;
Best Local Similarity 44.2%; Pred. No. 7.5e-07;
Matches 345; Conservative 0; Mismatches 436; Indels 0; Gaps 0;

QY 829 aatttgcgaaatcaaaagaaagaaagatttcattcattcattcattcattat 888
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15438 AAATTTTAAATGAAAGAAAAGAAAATGAAATATATAAAAAAATTTTATAAATAA 15497
QY 889 caaaccttcaaatctcatgttaattgaaagaaaataaagataaattgtttgtt 948
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15498 AAAAAAAGAAAAGGAGAAAAATTTTAAAAAATAAATAAATAAATA 15557
```

Qy	949	tttctctgtgtacattcttttcaggaaatcatattcctaacaacaaagtgtatttttaataatga	1008
Db	15558	TAAATTTTGTAGAAATAAAAAATGAAAGAAGATTATCAAAAAAATTTAAAAAATAATTTT	15617
Qy	1009	tgagagattaccatcccaaaaggctcgaaactataataacaagtttaataacttaacaataac	1068
Db	15618	ATATAAAAAAAATGATTTATAAAAAATATAAAAAATATAAAAAAGAGAAAAAACAATTA	15677
Qy	1069	atgtgatcaacaatcaatgcagagtttgcatttaaaatagaaatgattgagcaaacctcca	1128
Db	15678	AAAAAATAAATAATATATCATATAAAAAAACAACAAAAAAGAAAAAATAATATATAAATAAA	15737
Qy	1129	aaatgctctttaggatacacaagaatcttctcttagcttattaaagcggaggtccaact	1188
Db	15738	AAATATATATCATAAAAATATAAAAAATTTAAAAAATGTAAAAAATAAATATATACATAA	15797
Qy	1189	ctctctccctgtgagacttttggtttccaaatcttcttccaaacaaatccaataattag	1248
Db	15798	AATAAAAAATAATTTATTTAAATAAAAAAATAAATAAATAAATAAATAAATAAATAA	15857
Qy	1249	ttaatgggcataaatttggttttaaataagtcgaatgattttttagggaccatctctaat	1308
Db	15858	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	15917
Qy	1309	cacgcacaaatcatcctaattgtgaacacatttaaaacttaaaagtattgcatccacaatcc	1368
Db	15918	TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	15977
Qy	1369	ttaaaaatag	1428
Db	15978	CATATACATACAT	16037
Qy	1429	aaacgataaactcccttactcaacaattagcccaaaaaaacatccaatgcatcttaaaccta	1488
Db	16038	AT	16097
Qy	1489	ggaaatttaacaaactcaaataggttgtgagtttaaaaaaaacaaatagtagtgtaacat	1548
Db	16098	TAT	16157
Qy	1549	acgtacaccttaaaaaatatatactactcgaagaagtttttaaatcttttcgaaataataaca	1608
Db	16158	ACCTTAATATACAT	16217
Qy	1609	t	1609
Db	16218	t	16218

RESULT 4  
US-08-446-855A-1

; Sequence 1, Application US/08446855A

; Patent No. 5849573

; GENERAL INFORMATION:

APPLICANT: Stewart, Thomas S

APPLICANT: Flores, Maria V

; APPLICANT: O'Sullivan, William J

; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl

; TITLE OF INVENTION: phosphate synthetase II

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC

STREET: 1100 No. 5849573th Glebe Road, 8th Floor

CITY: Arlington

STATE: Virginia  
COUNTRY: USA

; COUNTRY: USA  
; 7TP: 22201-4

ZIP: 22201-4714  
COMPUTER READABLE

COMPUTER READABLE FORM:

MEDIUM LIFE: FLOPPY disk  
COMPUTER: IBM PC compatible

COMPUTER: IBM PC COMPACT  
OPERATING SYSTEM: PC-DOS

SOFTWARE: PatentIn Release

/ SOFTWARE. PATENTED RELEASE #1.24  
; CURRENT APPLICATION DATA:

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

```

, APPLICATION NUMBER: US/08/446,855A
, FILING DATE: 06-Jul-1995
, CLASSIFICATION: 435
, ATTORNEY/AGENT INFORMATION:
, NAME: Michard, Leonard C
, REGISTRATION NUMBER: 29,009
, REFERENCE/DOCKET NUMBER: 47-80
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 703-816-4000
, TELEFAX: 703-816-4100
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 8920 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: genomic
US-08-446-855A-1

```

Query Match 2.3%; Score 79.4; DB 2; Length 8920;  
Best Local Similarity 48.7%; Pred. NO. 3.5e-06;  
Matches 248; Conservative 0; Mismatches 256; Indels 5; Gaps 1;

Qy	1232	aaaaatcaataattgtaattggcgataataatttgggtttaaataagtcatagatttt	1291
Db	165	ATAAAAATAAAAAGATAAAAAATATTATATTGATATGTAGTATATAATAATGATAT	224
Qy	1292	ttaggaccatctcaatcacgcgaaatctccaaattgtgaacacatttaaaaactaaag	1351
Db	225	TCATATTATAACATAGATAAAAAACTTTTTTTTTTTTTTTCTTTATATTATAA	284
Qy	1352	tattgcattccaaatcccttaaatatataatatatatatatatatatatatatatat	1411
Db	285	CAATACATTTTAAGTTATTTTATATATATATATATATATATATATATATATATAT	344
Qy	1412	atatgaaagt-----Catatagaacgataaacctctactcaacaaattagcccaaaaaa	1466
Db	345	ATGTTGTGTGTTCAATTGTTTATAAAAATTACTTGAAATATAAACCTATTAAATATATT	404
Qy	1467	catcataatgcatttaaaactagggaattttaaacaactcaaatagggttgtagttaaana	1526
Db	405	CCAATTAAATAGAAATACAATTTATTAATTTTGTGTGTACACAAATTAATATAGTTTTACA	464
Qy	1527	aaacaaaatagtagatgcatcacatgcactcttcaaaaaatataactcatatcgaaagtttt	1586
Db	465	CTTCTTTATAAACCATCCATATATTATATACACAAATATATAATACTCCCAATATGT	524
Qy	1587	aaatttgcgaaattaaatcacatttatctatccaattaaataacattttaaataatgcataat	1646
Db	525	GGTTCCTATAAATTTTATATATATTTATTTATTTAAATTTATTTTCATTTATTTTTTTTT	584
Qy	1647	tcgtgaatccttatctttaaattccatatagaaacccaaacaaaaataaacatcatcaatag	1706
Db	585	CTTAGTTTATAAATAGTAATTTCTACTAATTTAAAAAATAAAAAAATAAAAAAATAAAA	644
Qy	1707	ttttaacttaacaaaaacgttagggaaaa	1735
Db	645	AGAAAAAATAAAAAATTTTACATATGAAAA	673

RESULT 5

US-09-150-741-1

03 03 130 741 1  
; Sequence 1, Application US/09150741 1

Patent No. 6183996

GENERAL INFORMATION:

; APPLICANT: Stewart et al.

; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl phosphate

; Patent No. 6183996

; TITLE OF INVENTION: Synthetase II

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/09/150,741

; CURRENT FILING DATE: 1998-09-10

```

; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match	2.3%;	Score 79.4;	DB 4;	Length 8920;
Best Local Similarity	48.7%;	Pred. No. 3.5e-06;		
Matches 248;	Conservative	0;	Mismatches 256;	Indels 5; Gaps
Qy 1232	aaaaaatoaaatagtgtaataatggcgataaattttggtttttaataagtcacatagatttt	1291		
Db 165	ataaaaaataaaagataaaaaatttatattgtagtgtagtataataatgattat	224		
Qy 1292	ttaggaccattcttaatocgcagaaatctctaaattgaaacacattttaaacttaaaag	1351		
Db 225	tcataattaacaatagataaaaaactttttttttttttttttttttttttatattat	284		
Qy 1352	tattgcattccaaatccttaaatatataatatataatatataatatataatatatat	1411		
Db 285	caatacatattgaagttatttatataatatataatatataatatataatatataatat	344		
Qy 1412	atatgaagt----tatataagaacgataaactccttactcaacaaattagcccaaaaaa	1466		
Db 345	atgttggtgttcattgtttataaaattcctgaaataaaaacttaataatatatt	404		
Qy 1467	catccataatgcattttaaactaggaaattttaacaaactcctaagtggtgtagttaaaa	1526		
Db 405	ccaatcaatatgaatacaaatatataattttgtgtgtgacacattaatatagttttaca	464		
Qy 1527	aaacaaaatagtagtgtagacatgcgtacctttaaaaatatatactcatatcgaaagttt	1586		
Db 465	cttcttataataaaacccatccatataattatacaacaatataataactcccacattgt	524		
Qy 1587	aaatttggaaaataaaacacatttatctatcaattaaaaacacattttaaataatgcataat	1646		
Db 525	ggttccataaattttattatataattttattttaatttttccatttatttttttttt	584		
Qy 1647	tctgtaattctcatctttaaatttccatatagaacccaaacaaaaataacatatacacaatag	1706		
Db 585	cttagtttataaaatagtaattctactaattttaaaaaaataaaaaaataaaaaa	644		
Qy 1707	ttttaacttaacaaaaacgttagggaaaa	1735		
Db 645	agaaaaaataaaattttacatatgaaaa	673		

RESULT 6  
 US-08-973-462-1  
 ; Sequence 1, Application US/08973462B  
 ; Patent No. 6191270  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DRUILHE, PIERRE  
 ; APPLICANT: DAUBERSIES, PIERRE  
 ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
 ; FILE REFERENCE: 0660-0125-0 PCT  
 ; CURRENT APPLICATION NUMBER: US/08/973,462B  
 ; CURRENT FILING DATE: 1998-02-06  
 ; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
 ; EARLIER FILING DATE: 1996-06-12  
 ; EARLIER APPLICATION NUMBER: FR 95/07007  
 ; EARLIER FILING DATE: 1995-06-13  
 ; NUMBER OF SEQ ID NOS: 29  
 ; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P.
US-08-973-462-1

```

Query Match 2.1%; Score 71.8; DB 4; Length 6152;  
Best Local Similarity 47.9%; Pred. No. 7.6e-05;  
Matches 280; Conservative 0; Mismatches 292; Indels 13; Gaps

[illegible]

RESULT 7  
US-07-638-431-1/c  
; Sequence 1, Application US/07638431  
; Patent No. 5198535  
; GENERAL INFORMATION:  
; APPLICANT: Hoffman, Stephen L.  
; APPLICANT: Charoenvit, Yupin  
; APPLICANT: Hedstrom, Richard  
; APPLICANT: Khumsmith, Srisin  
; APPLICANT: Rogers IV, William O.  
; TITLE OF INVENTION: Protective malaria sporozoite surface protein  
; TITLE OF INVENTION: immunogen and gene  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: A. David Spevack  
; STREET: NWRDC Building 1 T-12 National Naval  
; STREET: Medical Center  
; CITY: Bethesda  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20814-5044



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/00018  
FILING DATE: 19920103  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: Spevack, Avram D.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 295-6759  
TELEFAX: (301) 295-4033

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4673 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: N  
ANTI-SENSE: N  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium yoelii  
STRAIN: 17X(NL)  
DEVELOPMENTAL STAGE: erythrocytic stage  
TISSUE TYPE: Blood  
CELL TYPE: erythrocytic stage  
IMMEDIATE SOURCE:  
LIBRARY: Py-lambda gt11-2-7 kb genomic expression  
CLONE: Py10.1111  
FEATURE:

NAME/KEY: CDS  
LOCATION: 718..3195  
OTHER INFORMATION:  
PCT-US92-00018-1

Query Match 2.1%; Score 71; DB 5; Length 4673;  
Best Local Similarity 44.4%; Pred. No. 0.0001;  
Matches 530; Conservative 0; Mismatches 640; Indels 25; Gaps 5;

QY 510 agaaaaatcagacagacacagcaattatgtatttcttataaaagaaagaa 569  
DB 4563 AACGAACTAGAGGAGCAATATAAATTTATTCACATTATATGATCTTCAAAATTATGA 4504  
QY 570 attcatattactatagaattaaaagcgaagcgttgaaacgtgaaacgagaatttcta 629  
DB 4503 AAAGGTAGAGGTAATGCGCTTATTAGAATTTTATTCGATTTTATAAAGGAAATTA 4444  
QY 630 aaaaaaatagtaactgctacaaacttatttattgtgtatataacatatctataaagaac 689  
DB 4443 AAATAAATGATTTGATACAAATAAAGAATATATTTCCTTCTTAATAATATGAACAAGAC 4384  
QY 690 tcaaatatgataaatacttttaacaaatttctatgaattataataaaaaaagtcac 749  
DB 4383 AAAATGGATGAATATATATACAACTCAAAATTAATAAATTTACAAATAAAGAA-- 4324  
QY 750 ttttgacacttaaaaggttgacaaataacgctctcccaaaaaaaatcaaaacatttata 809  
DB 4325 ---TTAAGCAATAAAGAATTAATAAATTTATGCTGCCAACAAATAATGACACATGCA 4269  
QY 810 attctaaactatggtgtaattttgctgaaatcaaaagaaagaggaatttcttatac 869  
DB 4268 ACTACAAATAAATGAAGAAATGAATAAATTTGGAATAATGATATTCTCTCTTAATAAAT 4209  
QY 870 ataatgcttatttattatcaaaacttcaaatctgaatttgaagaaagaaataa 929  
DB 4208 AAAATTAATTTTAAATGATTTAAAAAATAAATAATCGATCTTATTTTGTAAAAATCAAA 1419  
QY 930 agataataatggttttttttcttattgttactatttctggaatataatattcatacaaa 989  
DB 4148 GAGAATAAATAATTAATAAATAATGCTCATATTTTTCAAAAACAGATACACGAATTT 4089

QY 990 aaaaatgtattttaaataatgatgatgagattaccatccaaaagggtogaacttat-----ata 1043  
DB 4088 GACAAATTTGCTCAAAATATGCTAAAAATATATCATCTTAAATTTTACACAAATAATATAATA 4029  
QY 1044 aaacaagtttaataactaaacaatacagtgatcacaatcaatgacagtttggatcttaaa 1103  
DB 4028 AATGAATCCAAACATAATAACACAGATATGAANAATCAAAATATGAATTTTCCGAAAGT 3969  
QY 1104 atagaaatgataggcaaacctcaaaaatgtctctcttaggatacacaataatcttcccttta 1163  
DB 3968 TTATATGAAAAATTAATAATTTTCAGATATGATAAATGAAGATGATAACTTCAAAAAAT 3909  
QY 1164 gattattaaagcgggaggttcaactctctccctctgtgagacttttctgttttcaaatctt 1223  
DB 3908 CATGAGCCAACTTTAGATATCTCTTTTAAACATCCCATCATTTTATTTTATACACTT 3849  
QY 1224 ttctttcaaaaaatcaataatagtggaagggcataatattgtttttaaataagtcac 1283  
DB 3848 TTAGTTCATAAACTTAAGACCATTTATTTTATGTTTGTAAATTTTAAATTAATACACA 3789  
QY 1284 tagatttttaggaccatctctaatacagacaaataatcctaattgttaacacatttataaa 1343  
DB 3788 TTTTGTGTTAT-----TTCCTTTTATCGATAATAATTGGTGGATTGCTCTATTTTGTAG 3735  
QY 1344 cttaaaagttatgcatcacaactccttaaaatatatatatatatatatatatatatat 1403  
DB 3734 GAATTCATTTTATATGATATATC-----ACTTTTGTGTTTATTCATAATATTTTGAA 3680  
QY 1404 atatatatatgaagtttatagaaacgataaactccttactcaacaatlagcccaaaa 1463  
DB 3679 AATAGTAATACCGTGTAAATATACAACTTAAATATGTTATTAATTTTAACTTTTAACTTTT 3620  
QY 1464 aaacatcccaatagtcatttaaaactaggaatttttaacaaactcaaaataggttgtagttaa 1523  
DB 3619 TTTTGTGTTATTTT-----TTTATATTAAGAATAATTTGTAACCAATTAATAATTTGAGTATA 3563  
QY 1524 aaaaaaataatagtagatgacacgacacgttcaactttaaataatatatactcatcgaagt 1583  
DB 3562 ATAAATATATATATTAACGAGACAAATTAGTTAAAAAATAATAGTTAAAAAATAATTCGT 3503  
QY 1584 tttaaatgttggaataataacatttatctatcaataataataacatttaataatgcat 1643  
DB 3502 TAAAAAATAATGAAATAAATTAATGATATATAATTCGAATCAATAAATAAAGATGACA 3443  
QY 1644 aattctgtatctatcttcttatttccatataagaaacaaacaaataaataa 1698  
DB 3442 ATTTATCAAACTGTTAATTTAAATAAATTAATCAATCAACAAAAAAGGAACAAA 3388

RESULT 9  
US-07-867-106-2/c  
Sequence 2, Application US/07867106  
Patent No. 5389526  
GENERAL INFORMATION:  
APPLICANT: Slade, Martin B  
APPLICANT: Chang, Andy C M  
APPLICANT: Williams, Keith L  
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular  
NUMBER OF INVENTIONS: Slime Moulds of the Genus Dictyostelium  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris  
STREET: One Liberty place 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106  
FILING DATE: 19920625  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PU PJ 7187  
APPLICATION NUMBER: PCT/AU90/00530  
FILING DATE: 02-NOV-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Feeney, Joanne Longo  
REGISTRATION NUMBER: 35,134  
REFERENCE/DOCKET NUMBER: RICE-0002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5852 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2378..5038  
NAME/KEY: CDS  
LOCATION: 2378..5038  
US-07-867-106-2

Query Match 2.1%; Score 70.6; DB 1; Length 5852;  
Best Local Similarity 42.8%; Pred. No. 0.00012;  
Matches 523; Conservative 0; Mismatches 689; Indels 9; Gaps 3;  
Qy 537 tttatgtattttcttataaaagaagaagaaattctatactattatagataaaagc 596  
Db 2443 TTTGAATTTTCCAAAAGTATTACAAAACCTTAAAAACCTATCCCAAGAAATAAGTTC 2384  
Qy 597 taagcagtgaaacgtgaagcagaattcttaaaaaaaatagtaaacctgcacaaactt 656  
Db 2383 GTCCATTTTAAATCTTTGAATAAATTAACCAAAACCTATTAAATTAATAAATAATAA 2324  
Qy 657 attatgtatataaacaatctataagaacactcaaatatataatgataaactatttaaca 716  
Db 2323 AATGAATTTACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2264  
Qy 717 aaattctatgaattataataaaaaagtcacitttgacacttaaaaggttgacaataa 776  
Db 2263 AATTAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAATATC 2204  
Qy 777 ccgtctctccaaaaaaatacaaacatttataatttctaaaaactatggtgtaattttgc 836  
Db 2203 AAAAAATAAATAAATAAATAAATTGTCATGCCAAACATGATAAATAATTGATATATATCC 2144  
Qy 837 tgaatcaaaaaagaagaagattctatcatcaatgaatttcattatgtatcaaaccttt 896  
Db 2143 AATATTATAAATAAGGTATTAATTAGATAGAGATAAATAAATAAATAAATAAATAAATA 2087  
Qy 897 caaatttcattgaattgaagaaaaaataaagataatgattgtttttttttttttttat 956  
Db 2086 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2027  
Qy 957 gttacattttcatggaatataattcatacaaaaaatgatttttaatatgatgagagat 1016  
Db 2026 TTTTATTTTTCAGAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1967  
Qy 1017 taccatcaaaaggtgcgaactatataaaaaaagaatttaataaactaaacatacatgtatc 1076  
Db 1966 GGTAAACATACATTAGTTTTTATAGTTTTTTCATATTTAAAAATAAATAAATAAATAA 1907  
Qy 1077 acaatcaatgacgtttttgtatcttaaatgagaatgattgagcaaacctcaaaaatgtct 1136  
Db 1906 ATTGATTTTAAATTATGAGATCTAATAAAAAAATAAATAAATAAATAAATAAATAA 1847

Qy 1137 tcttaggatcacaaatc---tttcccttagcttattaaagccggagttcaactctctc 1193  
Db 1846 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1787  
Qy 1194 tcccttgtagacttttggtttccaaatcttttcttccaaaaaatacaataattgattat 1253  
Db 1786 AAGTATATATCATGATAGGCAATTTATTTTATATATATCTATCTAAAAAATAAATAA 1727  
Qy 1254 gggcataatatttggttttaataagccatagatttttttttttttttttttttttt 1313  
Db 1726 AATGTCATCAAAATAGTATTTTAAACATTTTAAAAAATAAATAAATAAATAAATAA 1667  
Qy 1314 caaatatccctaaattgtaaacacatttaaaacttaaaagtattgattcacaactctaaa 1373  
Db 1666 CAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1607  
Qy 1374 atagaaacg 1433  
Db 1606 CTAGGTTAGTTTTTATAAATTTTACATATTTGTTAATAACTTTTAAATTTTGAATCAT 1547  
Qy 1434 ataactcttactcaacaatttagcccaaaaaaacatccataatgatttaaaactaggaa 1493  
Db 1546 GATATTACATCGTCCCGTTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1487  
Qy 1494 tttaacaaactcaaataggttggttagttaaaaaa---aacaaatagtagatgtacatac 1550  
Db 1486 TTTTAAAAAATGATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1427  
Qy 1551 gtacctttaaaatatatactcatcatcgaaagtttttaaaattttgcgaataataaatac 1610  
Db 1426 TTATCTTAATCAATTTTGGTTTATACATATTTATGTTCTGACTGAAGTATAGATCTT 1367  
Qy 1611 tatctatcaataaataacatttaataatgataattctgtaattctatctatcttatttc 1670  
Db 1366 ATTACTAAAGTTTCAAAAGTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATA 1307  
Qy 1671 catatagaaccaaaaaataaacaatatcaaatagtttttaacttaacaaaaacgttagg 1730  
Db 1306 TTTTTCATATCTCTCATGACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1247  
Qy 1731 gaaaagttgacctaaactagct 1751  
Db 1246 AAATTTATTTCTTCAATAGGT 1226

RESULT 10  
US-08-947-823-1/c  
Sequence 1, Application US/08947823  
Patent No. 6114605  
GENERAL INFORMATION:  
APPLICANT: Williamson, Valerie M.  
APPLICANT: Kaloshian, Isgouhi  
APPLICANT: Yaghoobi, Jafar  
APPLICANT: Bodeau, John  
APPLICANT: Milligan, Stephen  
TITLE OF INVENTION: Procedures and Materials for Conferring  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/947,823



; FILING DATE: 09-OCT-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/18802
; FILING DATE: 09-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,191
; FILING DATE: 10-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REFERENCE/DOCKET NUMBER: 023070-070210US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-947-823-1

Query Match 2.0%; Score 69.4; DB 4; Length 51952;
Best Local Similarity 43.4%; Pred. No. 0.00027;
Matches 523; Conservative 0; Mismatches 671; Indels 11; Gaps 4;

QY 471 atcctaaagtgttgacatttaccctgcttttccagaagaaatcagaacagaccac 530
DB 14613 ACCATAAGTCCTTTTAAATCAATATCTCTGCTTTTCAATATAAATAATATG 14554

QY 531 agcaaatattatgtatttcttataaaagaaagaaagaaattcatttacttatagaatt 590
DB 14553 ACCTGTAAATATATATTAATTAATCTAAAGAACGCAAGCCCAATTTTATAC 14494

QY 591 aaaaagctaaagcagtggaacgctgaagcagaatttctctaaaaataagtagtaaaactctac 650
DB 14493 AAAAAAAGTGGTGTTCATTTTCATCAACACACATTTGTAGACATTTGACATTTTCCACATTA 14434

QY 651 aaacttattatgtatataaactatctataaagaaactcaaatatataatgataaactcatt 710
DB 14433 AGAATTAATATATATATCAAAACATAAATCTTCAATAGGAGTATATAATATATTTT 14374

QY 711 ttaacaaatttctatgaattataaaagaaagtcacttttgacacttaaaaggttga 770
DB 14373 TTTACAAATAAATATATGTACTATATGTATATGTTCATATATAAGCGGAGAACTTTGCA 14314

QY 771 caataaccgtctctccaaaaaaatacaaaacatttataatttctaaaaactatggtgtaa 830
DB 14313 GAAAAATAAAC-----AACAGAGTTTAAATATGTACTCAAAACAAATAAATTAATATCA 14260

QY 831 ttttgcgaaatacaaaagaaagaggttctctatataagtttctatttctatttctatca 890
DB 14259 TAGTATATAATTAATGAAGAAAAAATCCCTCACAATATTAGAGCCCTATTCCTTGT 14200

QY 891 aacttcaaatctcatgtaatttgaagaaagaaataaagataatgttgtttttgtt 950
DB 14199 TTCAGAAACAGAGATAATTTGTCCTCAATGATCCATTTGAACAAAGATCTGGACTTATT 14140

QY 951 tcttattgtacatttctatggaataatattctatacaaaaaaatagtattttaaataatg 1010
DB 14139 ATTATTATTATAAGTCTTGTATTACATCGGAAGATTCCTAAAGAGATTAATTAATCA 14080

QY 1011 agagattaccatcaaaaggctgaacttataaaaaagagtttaataactaaacaatacat 1070
DB 14079 ACTGCCCTAACAAATGAATTAAT-TACTTTGAAGGCCCAATTAACAAAGAAATAAATAATATAT 14021

QY 1071 gtgatcacatcaatgacagtttcttctataaataagaaatgattgagcaaacctcaaaa 1130
DB 14020 ATATATACACATCTTGACAGTTTGTTCATTCATATAATATATACTAAATACAAATTCATTG 13961

QY 1131 atgtcttcttaggatcacaaaaatcttcttcttagctatttaaagccggaggttcaactct 1190
DB 13960 ACTTTTGAATTTATACATCTATCATGTTTAGGACAAACGCAATTTATCATATGTTACTTTT 13901

QY 1191 cctccctgtagaacttttttgcatttttcaaatcttttcttcttcaaa-aaatcaataaattag 1249
DB 13900 CTATGTTACAAGATTAAATGTCATTTAATTTTGTCTTAAACATGTTACGTGGAAAG 13841

QY 1250 taatggcataaatatttgggtttttaaataagtcocatttttttaggcacatctcctaatc 1309
DB 13840 TTAAGCTTAAATATATTGTCAAAAAAGGAAAGAGATTATTCTTTTAAACCAGATTAAAAA 13781

QY 1310 acgacaaaatctcctaaattgtaacacattttaaacttaaaagttaagcattcacaactcct 1369
DB 13780 TGAATTAAGGACATCTCTTTTAAATAAGAAATAATATATATATATATATATATATATAT 13721

QY 1370 taaaataga 1429
DB 13720 AT 13661

QY 1430 acgataaactccttactctcaacaaatagcccaaaaaaacatcccaatgcattttaaactag 1489
DB 13660 GAGGATG---CGCTTCACACGCTGTGTTTCATGTGTAATTTTATAGATACGTTAGAAATGAC 13604

QY 1490 gaatttcaacaaactcaaatagggttgtagttaaaaaaaacaaataagtagtagtgcacata 1549
DB 13603 CRAATCTCATGGAATATATATGTAATCGTAATGAATAATAAATGTAAACAAATTACAAA 13544

QY 1550 cgtaccttttaaaataataactcatcatcgaaagctttttaaatttgcgaaataaatacat 1609
DB 13543 ATTCAAGACTTACGGCAATATAAACATTCATAGAAAACTTAATTAAGAGTCTTAACAGAG 13484

QY 1610 ttatctatcaaatcaaatcaatttaaatgaatgaatcttctgtaattctcttcttcttctt 1669
DB 13483 TGAATAAAAAATTTAAAGAAACATAATATTAAAGAAAAAATAAATATATAGACATTATTC 13424

QY 1670 ccata 1674
DB 13423 TTATA 13419

RESULT 11
US-08-998-416-288/C
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 288:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: PAG1241RP  
US-08-998-416-288

Query Match 2.0%; Score 68.4; DB 4; Length 837;  
Best Local Similarity 48.1%; Pred. No. 0.00024;  
Matches 255; Conservative 0; Mismatches 271; Indels 4; Gaps 2;  
Qy 1199 ttagactttttgttttcaaatctttttccttcaaaaaataatagtttaattgggca 1258  
Db 694 TATATAGCTTATGGCTAAAGCAATTTGCTCTATAAACAATATATAGGTTGGATCCTTT 635  
Qy 1259 taattttggttttaattgaatccatagatttttttaggaccatctctaatcacgacaat 1318  
Db 634 TATAAGTATTTTAACTACATCTTTTATAATATTTTATTAAATTAATTAATGATATAA 575  
Qy 1319 atccataattgaacacatttaaaacttaaaagtattgcattcacatcccttaaaata 1378  
Db 574 ATATTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 515  
Qy 1379 tatatatatatatatatatatatatatgaaagtttatagaaagataac 1438  
Db 514 TTAATAATAATAAATAATAATAAAGAAATTAAGTTAAATTA-ATTTAAATAATAAT 456  
Qy 1439 tcttactcaacaattgcccac 1498  
Db 455 TCTTATAAAGAGATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 396  
Qy 1499 caaactcaaatagggtgtagttaaataaaacacacacacacacacacacacacacac 1558  
Db 395 TAAAAATAATAATTTACAAATTTTAAATAATAATAATAATAATAATAATAATAATAATA 336  
Qy 1559 aaaaataatactcatcgcgaagtttttaaattttgcgaatttaataacatttctctatc 1618  
Db 335 TTTTAAATAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 276  
Qy 1619 aataaaatacatttaataatgcataattctgtaatactctatctttaaattccatata 1678  
Db 275 AAAGAAAAATA---ATAATCTCAATAATATTATTAATACTAATTTAAATTTGAACATAG 219  
Qy 1679 accaaacacaaataacatacaaatagttttaaccttaacacacacacacacacacac 1728  
Db 218 ACTAATAGTAGTTCATATAATAATATTATTTTATAATAATAATAATAATAATAATA 169

RESULT 12  
US-08-998-416-186/c  
Sequence 186, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtie, Philipp

APPLICANT: Rebischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NO. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NO. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 2.0%; Score 67.6; DB 4; Length 615;  
Best Local Similarity 47.7%; Pred. No. 0.00032;  
Matches 232; Conservative 0; Mismatches 249; Indels 5; Gaps 1;  
Qy 1218 aatctttttcttcaaaaaatcaataattagtttaattggcacaataatttggttttaatta 1277  
Db 524 AATAATAATTTTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 465  
Qy 1278 agtcacatagatttttaggaccatctctaatcacgacacacacacacacacacacacac 1337  
Db 464 TAATAATAATTTCTATAAAGTTAAATAATAATAATAATAATAATAATAATAATAATAATA 405  
Qy 1338 ttaaaccttaaaagtattgcatcacaacacacacacacacacacacacacacacacacacac 1397  
Db 404 ATATTAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 345  
Qy 1398 atatatatatatatgaagttatagaaacacacacacacacacacacacacacacacacacac 1457  
Db 344 AATATATTTTATAAACAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 286  
Qy 1458 ccaaaaaac 1517  
Db 287 -TAATAATTTTATAAAGAAAAATAATAATAATAATAATAATAATAATAATAATAATAATA 230  
Qy 1518 agttcaaaaaac 1577  
Db 229 TTTGAACATAGACTAAATAGTATTTCATATAATAATAATAATAATAATAATAATAATAATA 170  
Qy 1578 gaaagtttttaatttgcgaatttaaacacacacacacacacacacacacacacacacacac 1637  
Db 169 AATAATGATGAATTAAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 110

Qy	1155	t t t c c t t a g c t t a a g a c g g g a g t t c a o a c t c t c t c c c t g t a g a c t t t t - - - -	1209
Db	638	T C T A C T T G G C A A A T A C A A T C C T G A G A T C A A A C C C G T G T T T T T A A T T T T T T T T T T	579
Qy	1210	t g t t t c a a a c t t t t t c t t c a a a a a c a a t a a t t a g t l a a t g g g c a a t a t t t g g t	1269
Db	578	T C T T T T T A C A T T T T A A A A A T T G A A T T A T A A A A T A T G A A A T T A T A A A A C T T T A A T	519
Qy	1270	t t t a a t t a a g t - - c c a t a g a t t t t t t a g a c c a c t c t c a a t c a a g a c a a a t c a t c c t a a a t	1327

```
Db 518 ATAAATATGTAATATAAATACCTTAAATATAAATATGTAATATAAATACCTTTATAA 459
Qy 1328 tgaacacatttaaaacttaaaagattgcattcccaactccttaaaatataatataat 1387
Db 458 AATATGTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 399
Qy 1388 atatatatatatatatatataatgaaagtttatataagaacgataaactccttactc 1447
Db 398 ACATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 339
Qy 1448 acaatttgcccaaaaacatccata-atgcatthaacttaggaattttaacaaactca 1506
Db 338 ACATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 279
Qy 1507 aataggttggttagttaaaaaaaacaaactagatgacatacgtacaccttaaaaaat 1566
Db 278 ACATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 219
Qy 1567 atactcatatgaaagtttttaaaatttgcgaatttaaaacattttatctatcaatataa 1626
Db 218 ATATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 159
Qy 1627 tacatttaataatgataaattctgtaataatctatctatctatctatctatctatct 1686
Db 158 ATATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 99
Qy 1687 aaaaataacatacaaatagttttaaacttaacaaacggttagggaaagttgacctaac 1746
Db 98 ATATTTTAATATAAATATGTAATATAAATATGTAATATAAATATGTAATATAA 39
Qy 1747 tagcttgattga 1758
Db 38 AATATTTAATTA 27
```

## RESULT 15

```
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36

Query Match 1.98; Score 64.8; DB 2; Length 665;
Best Local Similarity 49.6%; Pred. No. 0.001;
Matches 257; Conservative 0; Mismatches 247; Indels 14; Gaps 3;

Qy 598 aagcagttgaaacgtaaaacagaaatttctaaaaaaatagtaaaactgtcacaaactta 657
Db 4 AAGATGAGGTAAATGTTGTTTAAATATAATTAATTAATTAATTAATTAATTAATTAATTA 63
Qy 658 ttatgtgtatataacatatactataaaagaaactcaaaatataatgataaaatcatcttaacaa 717
Db 64 AATATTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 123
Qy 718 aattcttatgaaattataataaaaaaagtcacttttgacacttaaaagggttgacaataac 777
Db 124 AATATTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 183
Qy 778 cgtctctccaaaaaaacatacaaacatttataattctcaaaactatggcgaattttgct 837
Db 184 AATATTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 236
Qy 838 gaaatcaaaaaaagaaagagatttcttatataatgaatttcatttcatgtatcacaacttc 897
Db 237 TAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 296
Qy 898 aattctcatgtaatttgaaagaaaaaaataaagataataatgattgttttttttcttattg 957
Db 297 TAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 353
Qy 958 ttacattttcatggaatatataatcatacaaaaaaatgtattttaaatgatgatgagagatt 1017
Db 354 TTATAATTAATAATGTTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 413
Qy 1018 accatccaaaaggctgaacttatataaaacaaagtttaataaactaaacataacatgatgatca 1077
Db 414 ----TTATAATTAATAATGTTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 469
Qy 1078 caatcaatgacagtttgatctcttaaaatagaagaattgatt 1115
Db 470 TTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 507
```

Search completed: July 13, 2001, 00:22:31

Job time: 8179 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 22:21:17 ; Search time 218.86 Seconds

(without alignments)  
9823.335 Million cell updates/sec

Title: PCT-US01-13059-3

Perfect score: 3424

Sequence: 1 aagctaaagctaccacat.....aacggcatagctgactcagt 3424

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	259	7.6	936	22	AAF58252
C 2	259	7.6	936	22	AAF58254
C 3	259	7.6	936	22	AAF58257
C 4	259	7.6	936	22	AAF58259
C 5	259	7.6	936	22	AAF58262
C 6	259	7.6	936	22	AAF58255
7	252	7.4	936	22	AAF58252
8	252	7.4	936	22	AAF58254
9	252	7.4	936	22	AAF58257
10	252	7.4	936	22	AAF58259
11	252	7.4	936	22	AAF58262

12	252	7.4	938	22	AAF58255	Oligonucleotide D1
C 13	115.6	3.4	19124	18	AAT72882	Plasmodium var-7 p
C 14	115.6	3.4	19124	21	AAZ98287	Plasmodium var-7 p
15	96.4	2.8	3549	21	AAA70223	Plasmodium falcipa
16	95.2	2.8	8310	20	AAZ29911	cDNA encoding a SC
17	83.4	2.4	6644	20	AAZ33181	Base sequence of t
18	83.4	2.4	7372	20	AAZ33182	Base sequence of t
19	83.4	2.4	7797	20	AAZ33180	Cowpox virus bsr f
20	83.4	2.4	7996	20	AAZ33184	Base sequence of t
21	83.4	2.4	19124	18	AAT72882	Plasmodium var-7 g
22	83.4	2.4	19124	21	AAZ98287	Plasmodium var-7 p
C 23	81.2	2.4	4590	7	AAZ60472	Sequence encoding
24	80.2	2.3	244	22	AAF58238	Oligonucleotide D1
C 25	79.6	2.3	244	22	AAF58238	Oligonucleotide D1
26	79.4	2.3	8920	15	AAQ62924	Carbamoyl-phosphat
27	79	2.3	6124	11	AAQ03568	Sequence encoding
C 28	78.2	2.3	1864	8	AAZ71405	Sequence of ANS-1
29	77.2	2.3	6888	21	AAZ70114	Plasmodium falcipa
30	76.4	2.2	1864	8	AAZ71405	Sequence of ANS-1
31	76	2.2	15577	19	AAV35616	SHOX gene prelinin
32	76	2.2	32367	19	AAV35620	Human SHOX (short
33	75.6	2.2	1671	13	AAQ24134	50 kD subunit of S
C 34	75.2	2.2	8310	20	AAZ29911	cDNA encoding a SC
35	74.8	2.2	2503	15	AAQ53480	PNPX30 xylanase cD
36	74.8	2.2	15418	21	AAA63785	Nucleotide sequenc
C 37	74.8	2.2	15418	21	AAA63785	Nucleotide sequenc
38	74.6	2.2	1230	10	AAZ90223	Malaria-specific D
39	74.4	2.2	7620	21	AAZ70132	Plasmodium falcipa
C 40	72.4	2.1	5852	12	AAQ11710	Dictyostelium plas
41	72.4	2.1	9789	17	AAZ41852	cDNA encoding Plas
C 42	72.2	2.1	1864	15	AAQ78892	Aspergillus nidula
43	71.8	2.1	6152	18	AAZ78867	P. falciparum live
44	71.4	2.1	1431	21	AAZ37082	DNA sequence encod
C 45	71.4	2.1	2104	13	AAQ25273	DNA sequence encod

ALIGNMENTS

RESULT 1	AAF58252/c
ID	AAF58252 standard; DNA; 936 BP.
XX	
AC	AAF58252;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Oligonucleotide D1835.
XX	
KW	Electron-transfer group; ETW; mismatch; genotyping;
KW	gene expression; ss.
OS	Synthetic.
XX	
PN	WO200107665-A2.
XX	
PD	01-FEB-2001.
XX	
PF	26-JUL-2000; 2000WO-US20476.
XX	
PR	26-JUL-1999; 99US-0145695.
PR	17-MAR-2000; 2000US-0190259.
XX	
PA	(CLIN-) CLINICAL MICRO SENSORS INC.
XX	
PI	Umek RM;
XX	
DR	WPI; 2001-159728/16.
XX	
PT	Nucleic acids containing electron-transfer group, useful as labels in
PT	hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT	a single surface
XX	





Qy	625	ttctaaaaaaatagtaaacgtctacaacattatttggtgtatataacatatctataaa	684
Db	767	#####	728
Qy	685	gaaactcaaatatgatgaataatctttaaacaataattctatgaataataaaaaaa	744
Db	727	#####	668
Qy	745	gtcacttttgacacttaaaaggttgacaataacgctctccaaaaaaaatcaaaacat	804
Db	667	#####	608
Qy	805	ttaataattctaaaaactatggtgtaatttttgctgaaatcaaaaagaaagaggtttct	864
Db	607	#####	548
Qy	865	atatacaagtcttcattatgtatcaaaactttcaaatctcatgttaattgaaaggaaaaa	924
Db	547	#####	488
Oy	925	aattaagaataaataattttttgttcttatggtcacattttcattggaataatatattcat	984

[illegible][illegible]

RESULT 4  
AAF58259/C  
ID AAF58259 standard; DNA; 936 BP.  
XX  
AC AAF58259;







Pt		Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -	
Pt		(CLIN-) CLINICAL MICRO SENSORS INC.	
Pt		Umek RM;	
Pt		WPI; 2001-159728/16.	
Pt		The present invention relates to a composition comprising two nucleic acid each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, CC monitoring gene expression.	
Pt		Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;	
Pt		Query Match Best Local Similarity      7.4%; Score 252; DB 22; Length 936; Matches                  0.4%; Pred. NO. 4.8e-29; Mismatches                184; Indels     0; Gaps	
Qy	630	aataaaatagtaaacgtctacaacatttttggtgatatcaaatcatcctaagaagaac	689
Dd	1	www.....	60
Qy	690	tcaaataatgatgaactttaaccaaaatttcataatcataaaaaagtccac	749
Dd	61	www.....	120
Qy	750	tttgtaccttaaagggttagacaataaocgtotccccaaaaaaatacaaacattata	809
Dd	121	www.....	180
Qy	810	attctaaactatgggtgaatttgctgaaatcaaaagaagaaggatttcataatc	869
Dd	181	www.....	240
Qy	870	ataagtttcattatgtatccaactttcaaatctcatgtaattigaaaggaaaaatta	929
Dd	241	www.....	300
Qy	930	agataaaatgtgttttttgcatttatgattcaatttcattgaatatataatcacaa	989
Dd	301	www.....	360
Qy	990	aaaatgtattttaatgatgatgagattaccatcccuaaggltcgaaacttatataacaa	1049
Dd	361	www.....	420
Qy	1050	gttaataactaaacatatcatgtgatcacaatcaatgacagtittgatcttaaatagaa	1109
Dd	421	www.....	480
Qy	1110	atgattgagcaaacctcaaaatgtctcttaggatcacaaaatttccttttagcttat	1169
Dd	481	www.....	540
Qy	1170	taaaagccggagggtcaactctctccottgttagactttttgtttccaacttttctt	1229

[illegible]

```

Query Match      7.4%  Score 252;  DB 22;  Length 936;
Best Local Similarity 0.4%;  Pred. No. 4.8e-29;
Matches 3;  Conservative 599;  Mismatches 184;  Indels 0;  Gaps 0;

Qy 630 aaaaaatagtaaacgtcacaaattattatgtgtataaacatctataaagaac 689
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



D	b	121	atttcataaacatggtgaatttctgcagaaacaaaagaaggattttctatatc	869
Q	y	810	atttcataaacatggtgaatttctgcagaaacaaaagaaggattttctatatc	869
D	b	181	ataagtcttcatttgatcaaaccttcaaatttcocatgaatttgaaaggaaaaaat	929
Q	y	870	ataagtcttcatttgatcaaaccttcaaatttcocatgaatttgaaaggaaaaaat	929
D	b	241	agataataatgttgttttcttatgttacactttccatggaaatatataatcataca	989
Q	y	930	agataataatgttgttttcttatgttacactttccatggaaatatataatcataca	989
D	b	301	aaatgtattttaatatgatgagagattaccatccaagaagtcgaacctataaaaca	1049
Q	y	990	aaatgtattttaatatgatgagagattaccatccaagaagtcgaacctataaaaca	1049
D	b	361	ttaataactaaacaatacatgtgatcaccaatcaatgacagtttggatcttaaaataga	1109
Q	y	1050	ttaataactaaacaatacatgtgatcaccaatcaatgacagtttggatcttaaaataga	1109
D	b	421	atgattgagcaaacctcaaaaatgtctcttaggatcacaaaatcttcccttagcttat	1169
Q	y	1110	atgattgagcaaacctcaaaaatgtctcttaggatcacaaaatcttcccttagcttat	1169
D	b	481	taagccggagggtcaactctctctcccttgtagactttttgtttcgaatctttctt	1229
Q	y	1170	taagccggagggtcaactctctctcccttgtagactttttgtttcgaatctttctt	1229
D	b	541	tcataaaatcaataattagtttaatggcgataataattggttttaattgaagtcacatgatt	1289
Q	y	1230	tcataaaatcaataattagtttaatggcgataataattggttttaattgaagtcacatgatt	1289
D	b	601	ttttaggaccatctctaataacagacaaaatactcctaattgttaacacattttaaactaaa	1349
Q	y	1290	ttttaggaccatctctaataacagacaaaatactcctaattgttaacacattttaaactaaa	1349
D	b	661	agatgcatcacaactccttaaaaataataataataataataataataataataataata	1409
Q	y	1350	agatgcatcacaactccttaaaaataataataataataataataataataataataata	1409
D	b	721	atatat 1415	780
Q	y	1410	atatat 1415	780
D	b	781	www	786
RESULT 12				
ID	AAF58255			
XX	AAF58255 standard; DNA; 938 BP.			
AC	AAF58255;			
DT	24-APR-2001 (first entry)			
DE	Oligonucleotide D1876.			
XX	Electron-transfer group; ETW; mismatch; genotyping;			
KW	gene expression; ss.			
XX	Synthetic.			
XX	WO200107665-A2.			
PN	01-FEB-2001.			
PD	26-JUL-2000; 2000WO-US20476.			
XX	26-JUL-1999; 99US-0145695.			
PR	17-MAR-2000; 2000US-0190259.			
PA	(CLIN-) CLINICAL MICRO SENSORS INC.			

XX Umek RM;  
 XX WPI: 2001-159728/16.  
 DR  
 XX Nucleic acids containing electron-transfer group, useful as labels in  
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
 PT a single surface  
 XX  
 XX Example 6; Page 127; 159pp; English.  
 XX  
 CC The present invention relates to a composition comprising two nucleic  
 CC acids each containing an electron-transfer group (ETM) having  
 CC different redox potentials. The invention is used for electronic  
 CC detection of nucleic acids, especially of substitutions (mismatches)  
 CC and single-nucleotide polymorphisms, e.g. for genotyping,  
 CC monitoring gene expression.  
 XX  
 XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;  
 SQ  
 Query Match 7.4%; Score 252; DB 22; Length 938;  
 Best Local Similarity 0.4%; Pred. No. 4.8e-29;  
 Matches 3; Conservative 599; Mismatches 184; Indels 0; Gaps 0;  
 Qy 630 aaaaaatagtaaaactgctacaaacttatttattgtgtatataacacatctctataaagaac 689  
 Db 1 www. .... 60  
 Qy 690 tcaaatatgatataaacttttaacaaaatttctatgaattataataaaaaaagtcaac 749  
 Db 61 www. .... 120  
 Qy 750 ttitgacacttaaaagttgacataaccgtctctcacaataaaatacaaacatttata 809  
 Db 121 www. .... 180  
 Qy 810 attcttaaacatggtgttaattttgtgtaacaaagaagaaggatttctatc 869  
 Db 181 www. .... 240  
 Qy 870 ataagtttcattatgtatcaaaacttcaaaatttcattgaatttgaaaggaataaatta 929  
 Db 241 www. .... 300  
 Qy 930 agataatgtgttttttcttcttattacatttctcgaataatattcacaacaa 989  
 Db 301 www. .... 360  
 Qy 990 aaatgtattttaatgatgatgagattacacatccaaaggtcgaaacttataaaacaa 1049  
 Db 361 www. .... 420  
 Qy 1050 gtaataactaaacatacatgtgatcacaaatcaatgacagtttggatcttaaaatgaa 1109  
 Db 421 www. .... 480  
 Qy 1110 atgattgacaaacctcaaaatgtcttcttaggatcacaaatcttctttagcttat 1169  
 Db 481 www. .... 540  
 Qy 1170 taagcgggagttcaactctctcccttgtagacttttttttcaaatcttttctt 1229  
 Db 541 www. .... 600  
 Qy 1230 tcaaaaatcaataattagtttaaggcataaatttggtttttaatttaagtccaatgatt 1289  
 Db 601 www. .... 660  
 Qy 1290 ttttaggaccatcttaacacgacaaatatacctaaattgtgacacatttaaaacttaaa 1349  
 Db 661 www. .... 720

Qy 1350 agtattgcattcaacatcccttaaaatataatataatataatataatataatataat 1409  
 Db 721 www. .... 780  
 Qy 1410 atatat 1415  
 Db 781 www. ....  
 RESULT 13  
 AAT72882/c  
 ID AAT72882 standard; cDNA; 19124 BP.  
 XX  
 AC AAT72882;  
 DT 12-SEP-1997 (first entry)  
 DE Plasmodium var-7 gene.  
 XX  
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;  
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;  
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;  
 KW Plasmodium; ss.  
 XX  
 OS Plasmodium vivax.  
 OS Plasmodium falciparum.  
 PH Key Location/Qualifiers  
 FT exon 7317..15139  
 FT /\*tag= a  
 FT /number= 1  
 FT Intron 15140..16205  
 FT /\*tag= b  
 FT /number= 1  
 FT exon 16206..17552  
 FT /\*tag= c  
 FT /number= 2  
 FT /note= "no stop codon given"  
 WO9640766-A2.  
 PD 19-DEC-1996.  
 XX  
 PF 07-JUN-1996; 96WO-US09508.  
 XX  
 PR 07-JUN-1995; 95US-0487826.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;  
 PI Wellens TE;  
 XX  
 DR WPI: 1997-052231/05.  
 DR P-PSDB; AAW22475.  
 XX  
 PT New malaria vaccines - contains cysteine-rich DBL family protein  
 PT binding domains homologous domains of the Duffy and sialic acid  
 PT binding proteins  
 XX  
 PS Claim 4; Page 56-61; 96pp; English.  
 XX  
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to  
 CC the Duffy binding like (DBL) family of genes which have homology to the  
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein  
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The  
 CC var family of genes modulate cytoadherence and antigenic variation of  
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding  
 CC protein (DABP) are soluble proteins that appear in the culture  
 CC supernatant after infected erythrocytes release merozoites. DABP and  
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte  
 CC surface. These proteins are necessary for erythrocyte invasion by the  
 CC parasite. This sequence can be used in the compositions of the  
 CC invention. The compositions are for the treatment and prevention of



CC malaria, and comprise either a nucleotide sequence or encoded polypeptide  
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a  
CC family of genes having homology with conserved regions of DABP and SAPP.  
CC The compositions are used for the treatment and prevention of malaria.  
CC They are also used in the preparation of vaccines for inducing a  
CC protective immune response in a mammal to Plasmodium merozoites  
CC (especially Plasmodium falciparum or Plasmodium vivax).  
XX  
SQ

Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T; 0 other;

Query Match 3.4%; Score 115.6; DB 18; Length 19124;  
Best Local Similarity 43.3%; Pred. No. 4.1e-09;  
Matches 917; Conservative 0; Mismatches 1174; Indels 27; Gaps 7;  
QY 390 tatattgacctttcttactacttttttcagactatttgcctttttgacctcaaac 449  
DB 7083 TATATATATTAATAGTGTACTATTATAATATTAATATTAATATATGATATCTTAATAAATA 7024  
QY 450 ttttttgattgaccttttcgactcctaaagtggttgacaatttaacctgcttttctcca 509  
DB 7023 ATATTTAAATGATATATAATAATAATTAATAGTATTATATATTTTAAATAAATAATATA 6964  
QY 510 agaaaatcagacagacagcagcaaatattatgtatttctattaaagaaagaaaga 569  
DB 6963 TATATATTAATAAATTTATAATAATTTAAATATTTCTACAAAAAATAATATATCAGA 6904  
QY 570 attcatattacttagaattaaagcagtggaacgtgaaacgtgaaagcagaatttcta 629  
DB 6903 AATATTATATTTTATGATTCCTTTTATTTATCTAATTTAATATATATATTTT 6844  
QY 630 aaaaaatagtaaacgtgtacaacatttttggtgtatataacatctctataaagaac 689  
DB 6843 TATGTTTTTATTAAGTAAATTAATAATGAG-----AAAAAATAACGAAATAC 6789  
QY 690 tcaatatatgataatcatcttaacaaatttctatgaattataataaaaaagtcac 749  
DB 6788 AAACATATAAAGATATATGCAAGTGTTTATATATTTAATTTAATTAACATTATATA 6729  
QY 750 ttttgacacttaaaaggttgacaataaacgctctcccaaaaaaaatcaaacatttata 809  
DB 6728 TGTATATTTTTTTGACITTTATTTAAATTTATATATATATATATATATAGAGATAACA 6669  
QY 810 attctcaaacactagtgtaatttctgctgaacatacaaaagaaagagatttctcatc 869  
DB 6668 AAAGAGACAAATATGTTCTGTTCTCTCTATCTATATATTTATCTCATATTTATATAT 6609  
QY 870 ataagtttcattatgtatcaaaacttcaaaatttcatgtaatttgaaaggaagaaatta 929  
DB 6608 CATATATATAAATGATATAGATACATATTTCTTGTTGTTGTTGTTATATATAAAGTA 6549  
QY 930 agata-taagtggttttcttcttcttcttcttcttcttcttcttcttcttcttcttca 988  
DB 6548 GTATATATATATTAATTTTGTGTATATATATAATTTATATATATATATATATAT 6489  
QY 989 aaaaatgtattttaaataagtagagagattaccatcccaaaaggtcgcaacttatataaaca 1048  
DB 6488 ATAGCATCAAAAAAATAATGATAAATAAATAACAGGAAAAATATATATATATATAT 6429  
QY 1049 agttaactaaacaatacatatgtatcacaatcaatgacagtttttgatctttaaataaga 1108  
DB 6428 ATATATATTAATAAATGTTTTATCATATTTGTTGTTGTTGTTGTTTATGATATTTCA 6369  
QY 1109 aatgattgagcaaacctcaaaagtcttcttaggatcacaaaat--ctttccttagct 1166  
DB 6368 TCCATTTTATGATTTCAAAATTTTATGCTATAATATAAATAAATAAGTAAATAACACA 6309  
QY 1167 battaagccggagtgcaacctctctcccttgtagactttttgttttcaaatctttt 1226  
DB 6308 TTATAAATATATATATCAAAATGAGTATTAAATAAATGTTTCATGTTCTATATATTTA 6249  
QY 1227 ctttcaaaaaataaataatagtttaattgggcaataatatttgggttttaataagtcacatag 1286

DB 6248 TATAAATGAAATATATTTGTTATATATAATAATACATATATATGCTACTATATAATAATA 6189  
QY 1287 atttttaggaccatctcttaataacagacaataatctctaaattgttaacacattttaaactt 1346  
DB 6188 ATATCTTTAAAGTAGTATATACTA-----AAATATATAAAATGCAATGATAAAATAGTA 6136  
QY 1347 aaaaatgtgattcacaatcctctaaataatataatataatataatataatataatata 1406  
DB 6135 TAAATCATACATA 6076  
QY 1407 tatatatagaagttagaagacgataaactcctctcaacaatttagcccaaaaaa 1466  
DB 6075 TGCATATATAGTAGAATAATTTATTTATATTTCCAAATCTGATATTTGTTTATATTTGTT 6016  
QY 1467 catccataatgcattttaaacttaggaattttaaacaactcaaataggttggttagttaa 1526  
DB 6015 ATATTATAATAACAAAAAAGAACGACAAAGTAGTAAACAAAAAATAATAATAAATAA 5956  
QY 1527 aaaaacaatagtagtgatcacatgcgttaaccttttaaataatataactcatatcgaaagt 1586  
DB 5955 GGATGATATAATAAAACTAATTTACCTATTAGTATAGTTTAAATGTC---TTTATATATAT 5899  
QY 1587 aaatttgcgaataataacatacatcttctcatcaattaaataacatttaataatgcataat 1646  
DB 5898 ATATATATATATATTTTATACAAATATTTCTTCTGTAATATGCATATTTGTTAGTTATCTA 5839  
QY 1647 tctgtaatatctatctttaaattccatataagaccaaacaataaacaataatcaaaatag 1706  
DB 5838 TTTTATATATATATGATACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 5779  
QY 1707 ttttaacttaacaaaacggttaggaaaagttgacctgaactgctgattgagcgttgac 1766  
DB 5778 ATCTATCTTTAACACAGACATATACACAAA-----TAAATATATAACTTTATATGTAATC 5726  
QY 1767 ttgcaatgcgaagcgatatttccaatatatactacatagtagtattttattttaggaag 1826  
DB 5725 TATTCACATATAAATATATATAATGAACACCCCATATATACTTATAGCCTTATATACACCA 5666  
QY 1827 tttctaaaaaggttggtgagtggttacttctgttggaggatgctatttttccctcttg 1886  
DB 5665 ATCAAAACACATAAATAATATATATATAATAATATATATATATATATATATATATAGTAT 5606  
QY 1887 ccaataattttacagagtatgggataaactacatactcatgattatgaacgctcaacttta 1946  
DB 5605 TTTGTATGAATPACTATTATTT--TAACTAATTTTATTTATTTATATATATATATATAT 5548  
QY 1947 ttgaaaaacccctcaatacaccaaatatgctcactagattccaaaacgtagaccattgt 2006  
DB 5547 TTTTCAAAAACTAATCATGTTTAATATATATAATAATAATAATAATAATAATAATAA 5488  
QY 2007 atctaactcnaattctcaatcaagtagttaattaccgatgggtaagaaaagttaacoga 2066  
DB 5487 ATAAATTTTAAATATTTATATTTTATCTTAAATATTAATTTCTTATTTTCTATAATACAT 5428  
QY 2067 tataattatcaaaaagaagaataaagcaacagattcttaactcttatttttgggtatag 2126  
DB 5427 GTACGTATTTGTAATATATGTTTACATTTTAAATAATGTAATAATATGACTTATTTATTCAG 5368  
QY 2127 aacatttgtaaaaaactcctcaaaagatatgtaactgttttaaaataataaactcactgagat 2186  
DB 5367 TATTTCTTATTTTATTTTATTTTATTTTCCCTTATTAAGATTTAAATAATATTTTACTTAT 5308  
QY 2187 taattcttcagactcgtgttagctataataatgtcgaagttcttcttcttcttcttcttctt 2246  
DB 5307 TTGTTGTAAGAGATATATAAATATATATATATGATGAGTCCATAAATATATTAAGCTATTGGA 5248  
QY 2247 aacctaaagatatgtatatattttctgtgaattatgtatgataataatttgcattcatgta 2306  
DB 5247 TACAAAAACAGGATATTTATTTCTTATTTTCAATTTTAAATAAATAAATAAATAAATAA 5188  
QY 2307 caaacatttacttaaaaaactgatttttcttactcacaatgttgcactgaagcaaaaaata 2366  
DB 5187 TATAGACGATTCACATATATAAATAATTTAGTCCATATAATTTCAATAAATAAATAAATAA 5128



```
QY 1527 aaacaaatagtagtgcacacgtacacgttttaaaatatactcatatcgaagaatttt 1586
Db 5955 GGATGATATAATAAATAAATTAACCTTAGTATAGTATGTC--TTTATATATAT 5999
QY 1587 aaatttgggaattaaatcacatttctatcaattaaaatacatttaataatgcataat 1646
Db 5898 ATATATATATATTTTATACAAATATTTCTTGTGAATATGCATATTTAGTATCTTA 5939
QY 1647 tctgtaatactctatctttaaatttccatagaaaccaaataaacaatatcaaatag 1706
Db 5838 TTTTATATATATATGACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5779
QY 1707 tttaaacttaacaaagcttgagggaagttgacctaactgcttgattgacgttgaac 1766
Db 5778 ATCTATCTTTAACACACACATATACACAAA-----TAAATATATTAACTTATATGTCATC 5726
QY 1767 ttgtcaatgcgaagcgatatttcccaatatatactacatgtagtatttatttatggaag 1826
Db 5725 TATTACATATAAATAATATAAATGAACACACCCATATATCTTATAGCTTATATACACCCA 5666
QY 1827 ttctaaagaggttggtgaggttacttgggtggagtgctatttttctctcttg 1886
Db 5665 ATCAAAACACATATAATATATATATAAATATATATATATATATATATATATATATAT 5606
QY 1887 ccaataattttacgagtgatgggataactacatactactcatgattatgaaogctcaattta 1946
Db 5605 TTTGTATGAATACTATTATTT--TAACATAATTTATTTATTTATTTATATATATATATAAT 5548
QY 1947 ttgaaaaacctcctaatacaccaaatatgtcactagattcgaataacgtagacaaattgt 2006
Db 5547 TTTTCAAAAACAACTAATCAGTTTAATATATATATATATATATATATATATATATATAT 5488
QY 2007 atctaactcctcaaatcctcaatcaatgatttaatttcccgatggtaagaaaagttaaccga 2066
Db 5487 ATAATTTTAAATATTTATATTTATCTTAAATATTAATTTCTTATTTCTTATAATACAT 5428
QY 2067 tataataatacaaaagaagaataagtcacagattcttaactctcttatttttgggtatg 2126
Db 5427 GTTACGTATTTGTGAATATGTATGTACATTTTAAATAATGTATAATATGACTTTATTTACG 5368
QY 2127 aacattttacaaaaatctcaaaagatgtgaactgttttaaaatataataatctcagat 2186
Db 5367 TATTCCTTTATTTATTTATTTATTTTCCCTTATTAAGATTAAATAATATTTACTTTATTAC 5308
QY 2187 taattctcagactcgtgttagctataataatgtcaagagttcttctgttttcaaggaaa 2246
Db 5307 TTGCTTAAAGATATATATAATATATATATATGAGTCCCAATATATATTAAGCTATTGGAAC 5248
QY 2247 aaccttaagagatgtatatatttctctgaattatgatgataataattgctattctgtca 2306
Db 5247 TACAAAATCAGGATATTTATTTCTTATTTTCAATTTTAAATAAATAATTTATTTATTATTA 5188
QY 2307 caaacattcttcaaaaatcgtatttcttactactacaatgttgactagaacaaaaata 2366
Db 5187 TATAGACGCATTCACATATAAATAATATATAGTCCCAATATATTTCAATAAATAAATAAATA 5128
QY 2367 cattgattattgatcatcgtcaactgaattttcttcaggaggatataattctcaaacat 2426
Db 5127 CGAAATTTTCAATTTATGTAATGTAGTTAATTTATTAATAATGGACAAATTTTTTTTAAAT 5068
QY 2427 agcaagaatctcaataatgtttctcgactaccttttagacgaaatttttttaagattcg 2486
Db 5067 GTTATATAATAGGAACCTTTTAAAGATAAATAAATAAATAAATAAATAAATAAATAAATA 5008
QY 2487 taactgacttatggctct 2504
Db 5007 TACAGTAACATTTTTTTAT 4990
```

RESULT 15

AAA70223

ID AAA70223 standard; DNA; 3549 BP.

```
XX AAA70223;
AC
XX
XX DT 07-NOV-2000 (first entry)
XX
XX DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:356.
XX
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoasicide; infection; insecticide; ds.
XX
XX OS Plasmodium falciparum.
XX
XX PN WO200025728-A2.
XX
XX PD 11-MAY-2000.
XX
XX PF 05-NOV-1999; 99WO-US26796.
XX
XX PR 05-NOV-1998; 98US-0107131.
XX
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GARD/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the
XX PT diagnosis of P.falciparum infection -
XX
XX PS Disclosure; Page 545; 577pp; English.
XX
XX CC The present invention describes proteins and their fragments (I) encoded
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX CC vaccines against P. falciparum infection comprising (I) or (II).
XX CC (I) and (II) are useful for the development of vaccines against
XX CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX CC antibody raised to immunogens comprising the sequences of (I), are
XX CC useful in the detection of infection with P. falciparum. Furthermore,
XX CC (I) (especially when they are rifins or secreted or membrane proteins)
XX CC can aid the identification of drugs to treat or prevent P. falciparum
XX CC infection, or they can be used to identify drug resistance in
XX CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX CC subsequent identification of proteins encoded by it will help to expand
XX CC our understanding of parasite biology, a process hampered by the
XX CC complexity of the parasitic lifecycle, and provide new targets for
XX CC vaccine and drug development. Parasite resistance to drugs and mosquito
XX CC resistance to insecticides have led to a resurgence of malaria in many
XX CC parts of the world, and there is a pressing need for vaccines and new
XX CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX CC and protein sequences given in the present invention, but which are not
XX CC specifically mentioned within the specification.
XX
XX SQ Sequence 3549 BP; 1709 A; 352 C; 279 G; 1209 T; 0 other;

Query Match 2.8%; Score 96.4; DB 21; Length 3549;
Best Local Similarity 45.6%; Pred. No. 2.8e-06;
Matches 546; Conservative 0; Mismatches 631; Indels 21; Gaps 5;

QY 512 aaaaatcagaacgaccacagcaaatatttgatttttctattataaaaaaagaagaagaat 571
Db 1102 aaaaaaataataataataataataataataataataataataataataataataataa 1161
QY 572 tcattattactatagaattaaaagcagttgaaacgctgaaacgagaaagcttcttaa 631
Db 1162 ataattgaattatttaataataataataaagaaaaataattctattttttttta 1221
QY 632 aaaaatagtaaacgctacaaactttatttgcgtataaacatctatataagaagaactc 691
```

Search completed: July 13, 2001, 00:24:44  
Job time: 7407 sec

```
Db 1222 aaaaaaataataaaaaatgacgtaccataagtttaaaagagaaagatatgaac 1281
Qy 692 aaatatgatataaattcattttaacaaaaatttctatgaaattataataaaaaaagtcactt 751
Db 1282 acattataatgtgtgataaataataataaataccatcttattttttaataatttt 1341
Qy 752 ttgacacttaaaaggttgacaaataacogtctctcaaaaaaaatcaaaacatttataat 811
Db 1342 caagactctctattttataaaagtatatgaaaatttataaaaggcaacataataaat 1401
Qy 812 ttctaaactatgtgttaattttgtctgaaatcaaaagaaaggaaggttctctatatcat 871
Db 1402 tatttatgatgatcatgtttttttaagtcatttaagaaa-----tgtgttaaaaa 1454
Qy 872 aagtttcattatgtatcaaaactttcaaatttcatgttaatttgaaaggaaaaaataaag 931
Db 1455 gaattgtgcttatttcacagacagggatttgatatttatataaatggaaaacacatat 1514
Qy 932 atataatgtgtttttgtttcttattgttacattttcttggaatatatatattcatacaaaa 991
Db 1515 gaataatttgataataataatacaataataataataataaaaaataaacaataaatacaa 1574
Qy 992 aatgtatttttaatatgtgagagattaccatccaaaaggtgcgaacttatataaaacaagt 1051
Db 1575 tatgtataataaacagacaaggttaaaagataacaatgtcctatttccctctctc---t 1631
Qy 1052 taataactaaacaatacatgtgatcacaaatcaatgacagttttgtctttaaatagaaat 1111
Db 1632 tataaaagatgatatttttcacacattgaagattaccatttttcattcatataaaagatat 1691
Qy 1112 gattgagcaaacctcaaaaatgtctctttaggatacacaaaatcttctttagcttatta 1171
Db 1692 tatatatatgtttacaaaaaacaatgtgatgaataataaattattccataaaattattaa 1751
Qy 1172 aagccggaggttcaactctctc----tccctgtagactttttgttttcaaatcttttc 1227
Db 1752 ccatttaataacaataataaaacaaaataatgttctaaattatttagtcacaataattact 1811
Qy 1228 ttccaaaaatcaataatagttaattgggcataatatttggttttaatttaagtccataga 1287
Db 1812 ttgtatacaaaactaaattgtaaaactcagttaaagaaactgttattcactcttataaa 1871
Qy 1288 ttttttaggaccattctctaataacgacaaaatactctaataattgttaacacattttaaacctta 1347
Db 1872 taattatagaccattcttaaaaaaagaaataaaagaataataataataataataat 1931
Qy 1348 aagttattgtcacactccttaaaatatatatatatatatatatatatatatatat 1407
Db 1932 ctatttaaaaaataaaaaataataataataataataataataataataataataata 1991
Qy 1408 atatatgaagtatatagaacgataaactccttactcaacaaattagcccaaaaaaac 1467
Db 1992 atattattactatagtcaaaaaaaaaa---taagtgtgtaatatatacaacaacataa 2048
Qy 1468 atccataatgcatttaactaggaaattttaacaaactcaaatagggttggttaaaaaa 1527
Db 2049 tgttatgatgacatctaatacaaatattctatttagatcccttgagtgataaaagt 2108
Qy 1528 aaacaaatagtagatgtacatcgtacacttttaaaaaatatatactcatatcgaaagtttta 1587
Db 2109 acacaaattattattttataaaatttttaataaaaaagtaataatatataataatga 2168
Qy 1588 aattttgcgaataataatacatatttatcatcaattaaataacatttaataatgcataatt 1647
Db 2169 atggtccctttattttttattcttcttgattaaacagacacatgcttttataaaaaaagaag 2228
Qy 1648 ctgtaatatctatcttaattttccatatagaccacaaacaaataaacaatacaata 1705
Db 2229 gttctatatattatgttatattttt---catatacaaaaataatcataatata 2282
```

**THIS PAGE BLANK (USPTO)**

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 12, 2001, 21:53:12 ; Search time 4439.56 Seconds  
(without alignments)  
11929.464 Million cell updates/sec

Title: PCT-US01-13059-3  
Perfect score: 3424  
Sequence: 1 aagcttaaacactaccacat.....aacggcatagctgactcagt 3424

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues 2688314  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_ba3.\*  
4: gb\_in1.\*  
5: gb\_in2.\*  
6: gb\_in3.\*  
7: gb\_om.\*  
8: gb\_ov.\*  
9: gb\_pat1.\*  
10: gb\_pat2.\*  
11: gb\_ph.\*  
12: gb\_pl1.\*  
13: gb\_pl2.\*  
14: gb\_pl3.\*  
15: gb\_pl4.\*  
16: em\_bal.\*  
17: em\_ba2.\*  
18: em\_fun.\*  
19: em\_higo\_hum.\*  
20: em\_higo\_inv.\*  
21: em\_higo\_rod.\*  
22: em\_hig\_hum1.\*  
23: em\_hig\_hum2.\*  
24: em\_hig\_hum3.\*  
25: em\_hig\_hum4.\*  
26: em\_hig\_hum5.\*  
27: em\_hig\_hum6.\*  
28: em\_hig\_hum7.\*  
29: em\_hig\_hum8.\*  
30: em\_hig\_inv1.\*  
31: em\_hig\_inv2.\*  
32: em\_hig\_other.\*  
33: em\_hig\_rod.\*  
34: em\_hum1.\*  
35: em\_hum2.\*  
36: em\_hum3.\*  
37: em\_hum4.\*  
38: em\_hum5.\*  
39: em\_hum6.\*  
40: em\_hum7.\*  
41: em\_in.\*  
42: em\_om.\*  
43: em\_or.\*

44: em\_ov.\*  
45: em\_pat.\*  
46: em\_ph.\*  
47: em\_pl.\*  
48: em\_ro.\*  
49: em\_sts.\*  
50: em\_sy.\*  
51: em\_un.\*  
52: em\_vi.\*  
53: gb\_sts1.\*  
54: gb\_sts2.\*  
55: gb\_sts3.\*  
56: gb\_sy.\*  
57: gb\_un.\*  
58: gb\_vil.\*  
59: gb\_vil2.\*  
60: gb\_htg1.\*  
61: gb\_htg2.\*  
62: gb\_htg3.\*  
63: gb\_htg4.\*  
64: gb\_htg5.\*  
65: gb\_htg6.\*  
66: gb\_htg7.\*  
67: gb\_htg8.\*  
68: gb\_htg9.\*  
69: gb\_htg10.\*  
70: gb\_htg11.\*  
71: gb\_htg12.\*  
72: gb\_htg13.\*  
73: gb\_htg14.\*  
74: gb\_htg15.\*  
75: gb\_htg16.\*  
76: gb\_htg17.\*  
77: gb\_htg18.\*  
78: gb\_htg19.\*  
79: gb\_htg20.\*  
80: gb\_htg21.\*  
81: gb\_htg22.\*  
82: gb\_htg23.\*  
83: gb\_htg24.\*  
84: gb\_htg25.\*  
85: gb\_pr1.\*  
86: gb\_pr2.\*  
87: gb\_pr3.\*  
88: gb\_pr4.\*  
89: gb\_pr5.\*  
90: gb\_pr6.\*  
91: gb\_pr7.\*  
92: gb\_pr8.\*  
93: gb\_pr9.\*  
94: gb\_rol.\*  
95: gb\_rod.\*  
96: gb\_in4.\*  
97: gb\_pr10.\*  
98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3424	100.0	93045	14	ATT32M21
2	191	5.6	2382	13	AF367313
3	180.4	5.3	4601	6	DMU11584
4	180.4	5.3	19517	6	DMU37541
5	174.4	5.1	104992	60	AC005504
6	174.4	5.1	169546	60	AC004157
7	164.4	4.8	67970	96	PFMAL1P3
8	161.8	4.7	14867	4	AE001398

AL162875 Arabidops  
AF367313 Arabidops  
U11584 Drosophila  
U37541 Drosophila  
AC005504 Plasmodiu  
AC004157 Plasmodiu  
AL031746 Plasmodiu  
AE001398 Plasmodiu





4349..4432,4513..4501,4709..4823,4900..4823,4900..5026,5142..5281,5368..5429,5507..5573,5652..5783,5870..5947,6041..6151))	intron	/number=12
/gene="T32M21_20"	exon	complement(5948..6040)
/note="similarity to rngB protein, Dictyostelium		/number=12
discoideum, PIR:S68824"		complement(6041..6151)
/codon_start=1		/gene="T32M21_20"
/product="putative protein"	gene	/number=13
/protein_id="CAB85548.1"		complement(join(7298..7394,7484..7545,7615..7710,8136..8197,8761..8954,9380..9486,9592..9830,10135..10219))
/db_xref="GI:7406446"	exon	/gene="T32M21_30"
/translation="MEETRKIDVGDWHSNLAHDEWTPLVSGSRASARYKHAAVVVD	gene	complement(7298..7394)
EKLIVGSGRGVLDVQVFDLRLSLTWSSLLKLTKESSADNIOEDDGSLSUREAFPAL		/gene="T32M21_30"
SDHRMKGWKLILIGHSKSSDNMLVREFDLETHSCGVTDVGNVPASRGHSITL		/number=1
VGSRLVPGGEDKRRLLNDLHLHLERFDVETKOTRVPFRDHTAATHSRYLL		7298..10219
IFGCGSHSIFYSDLHILDLQTMWSQPHVQGVDPFRAGHAGITIDENWYIVGGDINS		/gene="T32M21_30"
TGCLTELNNKSLVSTHVEARHPLASEGLSVCSAFECENILAFPGYNGKYN	CDS	complement(join(7298..7394,7484..7545,7615..7710,8136..8197,8761..8954,9380..9486,9592..9830,10135..10219))
DIFVMRLKPGSSHPKIFKSPAAAAAASVTAAYAIKSDKSDYPPANPPLNGNS		/gene="T32M21_30"
LEPRDIRNIDTIKEEKRALESSTAETOVENAKLREKIDEVNSHTELSQLOSVQGO		/note="similarity to astrocytic NOVA-like RNA-binding
LISRSRCFKLEAQIAELQKALSGQSIEAEVEMLRQRQASDEEDGTVRQGSAGV		protein, Homo sapiens, U70477"
WGLFGR"		/codon_start=1
complement(3722..3810)	intron	/product="putative RNA-binding protein"
/number=1	exon	/protein_id="CAB85549.1"
complement(3811..3870)		/db_xref="GI:7406447"
/gene="T32M21_20"		/translation="WESTESYAAGSGPELAKRSPHPDSSADSAEKPTHIFLVNSA
/number=2	intron	AAGSVIGKGGSTITEFQAKSGARIQLSRNQEFFPCTTDRIIMISSIKEVVNGLELIL
complement(3871..3948)		
/number=2	exon	Query Match 100.0%; Score 3424; DB 14; Length 93045;
complement(3949..4269)		Best Local Similarity 100.0%; Pred. No. 0;
/gene="T32M21_20"		Matches 3424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/number=3	intron	
complement(4270..4348)		QY 1 aagcttaagctaccacaacatcgaaatttagtaaaagacccatgatttgaatttggaaattgt 60
/number=3	exon	
complement(4349..4432)		Db 62465 AAGCTTAAAGCTACCAACATCGAATTTAGTAAAGACCCATGATTTGAAATTTGGAAATTGT 62524
/gene="T32M21_20"		
/number=4	intron	QY 61 cggcaaaatcgagaagatagcgcacacggaacagtgaaacacccacaaagcgcgtaa 120
complement(4433..4512)		
/number=4	exon	Db 62525 CGGCAAAATCGAGAGATATAGCGCGACACGGGACAGTGAAGAACCAAGACGCGTAA 62584
complement(4513..4601)		
/gene="T32M21_20"		QY 121 gaatgaacagtgaggaggaagagaatcttaccgcatcttcgagggagaaagatgag 180
/number=5	intron	
complement(4602..4708)		Db 62585 GAATGAACAGTGGGAGAGAGAGAGAGAAATCTTACCGATCATTCGAGGAGAAAGATGG 62644
/number=5	exon	
complement(4709..4823)		QY 181 aatcagagaaaaatctggaaaaaaagaaattaaagagaaagagagagaaagtgaagg 240
/gene="T32M21_20"		
/number=6	intron	Db 62645 AATCAGAGAAAAATCTGGAAAAAAGAAATTAAGAGAGAGAGAGAGAGAGAGAGAG 62704
complement(4824..4899)		
/number=6	exon	QY 241 gaagatgcagtgagactgctatagccacatcccatggtgtgtgatgagagagagaga 300
complement(4900..5026)		
/gene="T32M21_20"		Db 62705 GAAGATGCAGTGAAGACTGCTATAGCCACATCCCACATGCTGTGTATGAGAGAGAGAGA 62764
/number=7	intron	
complement(5027..5141)		QY 301 gagaggttaagcagcaaaattgtggagagataaagagagagagagactgagcgagctcaag 360
/number=7	exon	
complement(5142..5281)		Db 62765 GAGAGTTAAAGCAGCAAAATTTGGAGAGATTAAGAGAGAGAGAGACTGAGCGAGTCAAG 62824
/gene="T32M21_20"		
/number=8	intron	QY 361 ttccgtcgtcgtgttttaaaagaaagaatccctatatattgctcttttttactacttattt 420
complement(5282..5367)		
/number=8	exon	Db 62825 TTCCGTCGTCGTGTAAAAAGAAAGAAATCCTATTATTGCTTTTCTTTACTACTTATT 62884
complement(5368..5429)		
/gene="T32M21_20"		QY 421 tcagactatttgcctattttgcctcaaaatttttgattgttcacttttcgactccctaaagt 480
/number=9	intron	
complement(5430..5506)		Db 62885 TCAGACTATTTCCTATTATTGCTTCAAACTTTTTTGATTTGTCACCTTTTCGATCCTCAAGT 62944
/number=9	exon	
complement(5507..5573)		QY 481 gttgacaattttacctgcctcttttctccaagaaaaatcagaacacagacacagcaaat 540
/gene="T32M21_20"		
/number=10	intron	Db 62945 GTTTCACAAATTTACCTGCCTTTTCTCCAAAGAAAAATCAGAACACACACACACAAATTA 63004
complement(5574..5651)		
/number=10	exon	QY 541 tgtattttctataaaaaagaaagaatcattacttataagaattaaaaagctaaag 600
complement(5652..5783)		
/gene="T32M21_20"		Db 63005 TGTATTTTCTATTAAAAAAGAAAGAAATTCATATTACTATTATAGAAATTAAGACTAAG 63064
/number=11	intron	
complement(5784..5869)		QY 601 cagttgaaaaagctgaagcagaattttctaaaaaaataagtaaaactgctacaaactattt 660
/number=11	exon	
complement(5870..5947)		Db 63065 CAGTTGAAACGCGTGAAGCAGAAATTTCTAAAAAATAAGTAAACGCTACAAACTTATT 63124
/gene="T32M21_20"		

Qy 661 atgtgtatataacatctataaagaaactcaaatatatatgataaatcatcttttaacaaat 720  
Db 63125 ATGTGTATATAACATATCTATAAGAGAACTCAATATATATGATATAATCATATTTTAACAAAT 63184  
Qy 721 ttctatgaaattataataaaaaagtcacttttgacacttaaaaggttgacaataacogt 780  
Db 63185 TTCTATGAAATATTAATAAAAAAGTCACCTTTTGACACTTAAAAAGGTTGACAATAACCGT 63244  
Qy 781 ctctccaaaaaaactcaaacattataattctcaaacatgtaatttttgctgaa 840  
Db 63245 CTCGCCAAAAAATCAAAACATTTATAATTTCTAAACATATGTTGTAATTTTCTGCGAA 63304  
Qy 841 atcaaaagaaagaggtttctcatatcataagtttcatttcttgatccaactttcaaa 900  
Db 63305 ATCAAAAGAAAGAGGATTTCTATATCATATAGTTTCATTATTGTATCAAACTTCAAA 63364  
Qy 901 ttctatgtaattgaagggaaaaaataagataaatgtgtttttgttttctctatgta 960  
Db 63365 TTTTCATGTAATTTGAAAGGAAAAAATTAAGATATAATGTTGTTTGTGTTTCTTTAAGTTA 63424  
Qy 961 catctcatgaatatattcatacaaaaaaagttaataactaacaatacatgtagagattacc 1020  
Db 63425 CATTTTCATGGAATATATTTCTAACAACAAAAATGTATTTTAATGATGAGAGATTACC 63484  
Qy 1021 atccaaaggttcgaacttatataaaacaagttaataactaacaatacatgtagacaa 1080  
Db 63485 ATCCAAAGGTGCAACTTATATAAAACAAGTTAATACTAAACAATACATGTGATCAAA 63544  
Qy 1081 tcaatgacagtttgatcttaaatagaaatgtagtgagcaaccccaaaatgtctcttt 1140  
Db 63545 TCAATGACAGTTTGTATCTTAAAAATAGAAATGATGTGCAAACTCAAAAATGTCTCTTT 63604  
Qy 1141 aggatcacaaactctctttagcttataaagccggaggttcaactctctcccttg 1200  
Db 63605 AGGATCACAAATCTTCTTACCTTATTAACCGGAGTTCAACTCTCTCTCCCTTG 63664  
Qy 1201 tagactttttgttccaacttttcttccaataatcaataatagtttaagtgggata 1260  
Db 63665 TAGACTTTTGTCTTCAAACTTTTCTTCAAAAATCAATAATTAGTTAATGGGCATA 63724  
Qy 1261 atatttggtttaatgaagtcocatagatttttttagggaccatctctaatacagacaatat 1320  
Db 63725 ATATTTGGTTTTTAATTAAGTCCATAGATTTTTTATAGGCCACTCTCTAATCAGCAAAATAT 63784  
Qy 1321 cctaaattgtaacacatttaaaacttaaaagttatgtcattcacaaactcttaaaatatata 1380  
Db 63785 CCTAAATGTAACACATTTAAACTTAAAGTATTGCAATTCACAACTCCTTAAATATATA 63844  
Qy 1381 tatatatatatatatatatatatatatataatgaaagttaataagaaacgataactc 1440  
Db 63845 TATATATATATATATATATATATATATATATGAAAGTTATAGAAACGATAACTC 63904  
Qy 1441 cttaactcaaatagccccaaaaaacatccaatgaatgatttaactaggaattttaaca 1500  
Db 63905 CTTACTCAACAAATTAGCCCCAAAAAATCCATTAATGCATTTAAACTTAGGAATTTTAAACA 63964  
Qy 1501 aactcaaataggttggttagttcaaaaaaaacaaatagtagatgtacatacgtacacttaa 1560  
Db 63965 AACTCAAAATAGTTGGTAGTTTAAAAAACAATACTAGATGTACATACGTACCTTTAA 64024  
Qy 1561 aaatataactcatatcgaaagttttaaattttgcgaaattaaataacattatctatcaa 1620  
Db 64025 AAATATATACTCATATCGAAAGTTTAAATTTTGCAGAAATTAATAATACATTTATCTATCAA 64084  
Qy 1621 ttaaaatcattttaaatgcataattctgataatctatctttaaattttccatataaac 1680  
Db 64085 TTAATAATACATTTAATAATGCAATTTCTGTAATATCTATCTTTAAATTTCCATATAGAAC 64144  
Qy 1681 caaaacaaaataacacatatcaaatagtttttaacttaacaaaaaacggttaggaaaaagtta 1740  
Db 64145 CAAAAACAAATAAACATATCAATAGTTTAACTTAACTAAACAAAGGTTAGGAAAAAGTTGA 64204

Qy 1741 cctaactagctgtgattgacgttgaaactgtcaatcgaaagcgatattttccaatatatac 1800  
Db 64205 CCTAACTAGCTTGTGATTGACGTGAACTTGCAATGCGAAAGCGATATTTCCAATATATAC 64264  
Qy 1801 tacaatgtagtattatttatatttgaaagtttctcaaaaaggtgttgagtgattgttacttgt 1860  
Db 64265 TACATGTAGTATTAATTTATATGGAAGTTTCTAAAAAGGTGTGAGTGGATTTGTTACTTGT 64324  
Qy 1861 tggaggtgctattttctctctctgcataaatttttacgagtatgggataaactacata 1920  
Db 64325 TGGAGGATGCTATTTTTTCTCTTCCATTAATTAATTTACGAGTATGGGATTAACATATA 64384  
Qy 1921 ctcatgattatgaacgcctcaactttatttgaaaacccctctcaatacaccacaaatgtcac 1980  
Db 64385 CTCATGATTATGAACGCTCACTTTATTGAAAAACCTCTCTAATACACCAATATGTACAC 64444  
Qy 1981 tagattccaaaacgtagaccaaattgtatctaaatctcaaaatctcaatacaaaagttaaat 2040  
Db 64445 TAGATTCCAAAACGTAGACCAAAATTTGTTATCTAACTCAAAATCTCAATCAAAAGTATTAAT 64504  
Qy 2041 taccgatggttaagaaagtttaaccgataaatttatcaaaagaaagaaataagtcacacagat 2100  
Db 64505 TACCGATGGTAAGAAAAAGTTAACCGATATTAATTAACAAAAGAAAGATTAAGTCAACAGAT 64564  
Qy 2101 tcttaactctctttatttttgggtatatgaacattttgtacaaaactctcaaaagatatgaac 2160  
Db 64565 TCTTAATCTCTTTTATTTTGGTATATGAACATTTGTCAAAAAATCTCAAAAGATATGTAAAC 64624  
Qy 2161 tgtttaaaaataaaattcaactgagatttaattcttcagactcgtgttagctataataatgt 2220  
Db 64625 TGTTTAAAAATATAAATTTCACTGAGATTAATTTCTTTCAGACTCGTGTAGCTATATAATATGT 64684  
Qy 2221 caagagttctctctgttccaagaaaaaaccttaagatatgtatatatttctgtaattatg 2280  
Db 64685 CAAGAGTCTCTCTGTTTCAAGGAAAAAACCTTAAGATATGTATATTTTCTGTATTAATATG 64744  
Qy 2281 atgataaatttgcatttctattgttcacaaacattcaactttaaaaaactgattttcattac 2340  
Db 64745 ATGATATAAATTTGCTATTTCATTTGTCACAAACATTTACTTTAAAAAATCGTATTTTCATTAC 64804  
Qy 2341 tacaattgtgactaagaacaaaaataacattgatttattgatatatcgtcaactgaatttcc 2400  
Db 64805 TACAATGTTGACTAAGAACAAAAATACATTGATTTATGATATATCGCAACTGAATTTTC 64864  
Qy 2401 ttccgaggggataaattctcaacacatagcaagaatctcataaataattgttcgtgactacc 2460  
Db 64865 TTCCGAGGGATATAATTTCTCAACATAGCAAGATCTCATATAATATGTTTCGTGACTATCC 64924  
Qy 2461 tttagacgaaatttttttaagattcgttaacgtgacttatggtctctctgctgtgggggtca 2520  
Db 64925 TTTTAGCGAAATTTTAAAGATTTCGTAAGCTGACTTATGCTCTCTTCTGCTGTGGGGTCA 64984  
Qy 2521 atgcgaataaatctaaatgtatgggagtcacaaataaaataccaagaaaaataaaggagcag 2580  
Db 64985 ATGCGAATAAATCTAAATGTATGGAGTCAAAATAAAATACCAAGAAAAATAAAGGAGCAG 65044  
Qy 2581 caccataaactatatagggaccagaaactcttctcatgtgttttaaaatagagattatcccc 2640  
Db 65045 CACCCAATAAATATATATGGGACCAGAAATCTTTTCATTGTTTAAAAATAGGATTAATCCCG 65104  
Qy 2641 aagagtagaaggactaaattgaactgattggggtaggaagagacatccgtcacaaatcata 2700  
Db 65105 AAAGATGAAGGACTAAATTTGAAATGATTTGGGGTAGGAAGAGATCGGTGCACAAATCATTA 65164  
Qy 2701 atggctccacgcggaactgtcgtttatcaatttcatttaacttcggtcgcggttta 2760  
Db 65165 ATGGCTTCCACGGGAAACCTTGCGTTTATACAAATTTTCATTAATTTTCGGGTTCGGGTTTA 65224  
Qy 2761 tattccaaatgggtcacaataatattagttttaatacactaacgcgagtaataatttgggtgac 2820  
Db 65225 TATTCCAAATGGGTCAAAATTAATTAAGTTTAACTAACACTAACGGAGTAATTAATTAATTTGGTGAC 65284  
Qy 2821 tacaattttatcagtttgggtgcgaattagaaacgaaacatagtcgtaaataacgaggttcggt 2880



```
RESULT 3
DMU11584
LOCUS DMU11584 4601 bp DNA INV 23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION U11584
VERSION U11584.1 GI:508826
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.
SOURCE fruit fly.
ORGANISM Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 4601)
Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
Sequence, Organization and Evolution of the A+T Region of
Drosophila melanogaster Mitochondrial DNA
Mol. Biol. Evol. 11, 523-538 (1994)
94285822
2 (bases 1 to 4601)
Kaguni,L.S.
Direct Submission
TITLE Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of
JOURNAL Biochemistry, Michigan State University, East Lansing, MI,
48824-1318, USA
FEATURES
source
1..4601
Location/Qualifiers
/organism="Drosophila melanogaster"
/organella="mitochondrion"
/strain="Oregon-R"
/db_xref="taxon:7227"
/dev_stage="embryo"
650..1022
/note="repeat I-A"
/rpt_type=tandem
1023..1360
/note="repeat I-B1"
/rpt_type=tandem
1361..1705
/note="repeat I-C/A"
/rpt_type=tandem
1706..2043
/note="repeat I-B2"
/rpt_type=tandem
2044..2388
/note="repeat I-C"
/rpt_type=tandem
2491..2511
/note="deoxythymidylate stretch"
2512..2648
/partial
/rpt_type=tandem
2649..3112
/note="repeat II-A"
/rpt_type=tandem
3113..3576
/note="repeat II-B1"
/rpt_type=tandem
3577..4040
/note="repeat II-B2"
/rpt_type=tandem
4041..4504
/note="repeat II-C"
/rpt_type=tandem
complement(4565..4585)
/note="deoxythymidylate stretch"
BASE COUNT 2271 a 131 c 74 g 2125 t
ORIGIN
Query Match 5.3%; Score 180.4; DB 6; Length 4601;
Best Local Similarity 44.0%; Pred. No. 1.le-14;
Matches 1296; Conservative 0; Mismatches 1611; Indels 39; Gaps 11;
```

```
Qy 390 tatattgccttttcttactacttttattttcagactatttgccttattttgcctcaaac 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1672 TATATATATATATATATATATATATATATATATATATATATATATATATATATAA 1731
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 450 ttttttgatgtcacttttcgatcctctaaagtggttgacaatttaccgcctttttctcca 509
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1732 TAATTTATTTTAATCAGTAAATCTGAAATTAATTAATTAATTAATTAATTAATTAAT 1791
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 510 agaaaaatcagaacagaccagcaaatattatgtattttcttataaaaaaagaagaaga 569
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1792 AAAATGAAATAAATTTATTTCCCTCATTAATAATTTATTTGTTATTAATAACTTAANA 1851
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 570 attcatattacttagaattaaaaagctgaagcagttgaaaaacgtagaagcagaatttcta 629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1852 AATATTTTTTTTTTAAAAAAAATGATTTATTAATTAATTAATTAATTAATTAATTAAT 1911
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 630 aaaaaaatagtaaacctgctcaaaacttatttatgtgtatataacacatatctataagaagaac 689
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1912 ATAAATTTATTTATAAATAAATTTATTTAAATAATTAATTAATAAATAATTTTAAATATAAT 1971
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 690 tcaatatatgataaatacattttaacaaaaatttctatgaataataataaaaaaagtcac 749
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1972 AAAAATTTAAATGATTTTTTATAAAAAATTCAAATTCATATATATATATATATATACATAT 2031
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 750 tttgacaccttaaaaggttgacataaacctctctcccaaaaaaaatcaaacacatttata 809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2032 AATTTAAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2091
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 810 atttctaaaaactatgggtgtaatttttgctgaaatcaaaagaaagaaggatttctataac 869
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2092 CTGAATTAATTAATTTGTTATATATATATATATATATATATATATATATATATATATATAT 2151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 870 ataagtttcatattgtatcaaacctttcaaatttcatgtaatttgaaaggaaagaaata 929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2152 ATCCCCCTATTCATAAATTTATTTATATATTAATTAATAACTTAAAAAATATTTTTTTTAAA 2211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 930 agatataatgtgtttttgttcttctatgtgtacatttctcatggaatatatatcacaaca 989
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2212 AAAAAATTTATTTATAAATTAATTAATTAATAAATTTTATAAATAAATTTTATAAATAAATTT 2271
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 990 aaaaatgtattttaatatgatgagagattaccatccaaaggctgaacttatataaacaacaa 1049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2272 TAAAAATTTATTTAAAAAATTAATTAATAAATTTATATATATATATATATATATATATATATAT 2331
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1050 gttataactaaacaacatcatgtgatcacatcaatgacagtttttgatcttcaaatagaa 1109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2332 GAAAAATAATTTTAAATTTTAAATAAT - AAATAAATTTTAAATAATTAATTAATAAATAA 2389
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1110 atgattgagcaaacctcaaaaatgtctctcttaggatcacaaaaatcttctcctttagcttat 1169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2390 TCTATTTCATTTAATTAATTTAATTAATAAATAAATAAATTTAATAACTAATAATTAATAATAA 2449
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1170 taaagccgggagttcaactctctccctgttagactttttgttttccaaatcttttctt 1229
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2450 AATTTATTTTACTAATATTTAAATTAATAAATAAATAAATTTTATTTTATTTTATTTTATTTT 2509
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1230 tcaaaaaatcaataattagtttaaggcgcaataatttggttttttaagtccaatagatt 1289
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2510 TTAATAATTTTAAATTTATTTATATATATATATATATATATATATATATATATATATATATAT 2569
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1290 ttttaggaccatctctaatacgcgcaaaatcctcaaatgttgacacattttaaaacttaaa 1349
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2570 ATATATATATATATAGAAAAAATAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAA 2629
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1350 agatgtcattccaacatccctaaataatatatatatatatatatatatatatatatatatat 1409
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2630 TTTCTTAAATGTATTTTATATATATATATATATATATATATATATATATATATATATATATATAT 2689
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1410 atatatgaaggttatatagaacgataaactccttactcaacaattagcccaaaaaaacat 1469
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2690 AATAAACAAAAAATTTTATTAATAAATAAATTTTATTAATAAATAAATAAATAAATAAATAAATAA 2740
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



MEDLINE REFERENCE	83090428		/product="NADH dehydrogenase subunit 2"
AUTHORS	2 (bases 5269 to 5595)		/protein_id="AAC47811.1"
TITLE	Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R. Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes		/db_xref="GI:1166530"
JOURNAL	Nucleic Acids Res. 11 (8), 2411-2425 (1983)		/translation="MFNNSSKILFITITIMIIGTLTIVTSNWLGAMGLEINLSFTPLP LSDNNMLSTEASLKYFLTOVLASTVLFSSILLMLKNMNMNEINESFTSMIIMSALL LKCAAPHEWEPNNMEGLTFMNALMTWKQTAPLMLISYLNKYLLISLVLSVII CAIGLNQTSLRKLMAFSSINHLCWMLESSIMISESWLIILFFFYFSVLVFMENIF KLHFNLQIFSVMNSKILKFTLFEMNFLSLGGLPPFLGFDPKRLVLTIOOLTLCNOYFMLT IMMSTLTIITFFYLIRICYSAFMMNYFENNIMKMNMNSINYNMYMITMTFFSIFGLFLI SLTYFMF"
MEDLINE	83220794	tRNA	1264..1329
AUTHORS	3 (bases 404 to 5272) de Bruijn,M.H.		/product="tRNA-Trp"
TITLE	Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code	tRNA	complement(1322)..1383)
JOURNAL	Nature 304 (5923), 234-241 (1983)		/product="tRNA-Cys"
MEDLINE	83245048	tRNA	complement(1403..1468)
AUTHORS	4 (bases 804 to 1778) Satta,Y., Ishiwa,H. and Chigusa,S.I.	CDS	join(1470..1472,1474..3009)
TITLE	Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species		/note="mechanism underlying reading frame shift after first codon uncertain"
JOURNAL	Mol. Biol. Evol. 4 (6), 638-650 (1987)		/codon_start=1
MEDLINE	88174373		/transl_table=5
AUTHORS	5 (bases 5268 to 13619) Garcesse,R.		/product="cytochrome c oxidase subunit I"
TITLE	Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations		/protein_id="AAC47812.2"
JOURNAL	Genetics 118 (4), 649-663 (1988)		/db_xref="GI:7412849"
MEDLINE	88212147		/translation="MSRQLWFSTNHDKDIGTYPIFGAWAGVGCTSLILIRAEIGHGP ALIGDOIYNVIVTAHAFAIMEFWMPIMIGFGNMLVPLMLGAPDMAFRMNNKSFV LLPPALSLLSVSNVENGACTGTVPPLSAGIAHGASVDLAIFSLHLAIGSSILGA VNFVTINMRSTGISLDRMPLFVWSVITALLULLSLPVLAGAITMLLTDNRNLTSF FDPAGGGDPILYQHLEWFFGHPEVIYILIPFCGMISHIIISQSGKKETFGSLGMTIYA LAIGLGFIVMAHHMTVGDMDVTRAYFVSATMIIAVPTGIKFISWLATLHGQTLSYS PAILWALGFVLFTVGGTGVLANSSVDIILHDTVVVAHPHYVLSMGAVFAIMAGF IHWPFLTGLTNANKLKSHFIIMEIGVNLTFFPOHFLGAGMPRYSDDYPDAYTTWN IVSYTGISTLSLGLIFFEFFIIESWSQRSQRYVIPQIOLNSSIEWQNTTPPAHSYSELP LLTN"
AUTHORS	6 (bases 441 to 2967) Satta,Y. and Takahata,N.	tRNA	3012..3077
TITLE	Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup	CDS	/product="tRNA-Leu"
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)		3083..3767
MEDLINE	91088557		/note="TAA stop codon is completed by the addition of 3' A residues to the mRNA"
AUTHORS	Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.		/codon_start=1
TITLE	Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods		/transl_except=(pos:3767,aa:TERM)
JOURNAL	Science 258 (5086), 1345-1348 (1992)		/transl_table=5
MEDLINE	93088057		/product="cytochrome c oxidase subunit II"
AUTHORS	8 (bases 14917 to 19517) Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.		/protein_id="AAC47813.1"
TITLE	Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA		/db_xref="GI:1166532"
JOURNAL	Mol. Biol. Evol. 11 (3), 523-538 (1994)		/translation="MSTWANLGQDSAPLMEOILFFDHALLILVITLVXYLWEM LFNNYVRNRELLHQLIEMTWITLPAIILLFIALPSRLLYLLDEINEPSVTKSIGH QWYWSYEIDFNIEFDYSMTNELMDTGFLEDNDNRVLPNMNQIRLVITAADYI HSWTPALGVKVVDGTPGRNLQNTFFINRPGFLFYGCQCEICGANHSMPPIVIBSPVNPY FIKWISSNNS"
MEDLINE	94285822	tRNA	3768..3838
AUTHORS	9 (bases 1 to 408; 13319 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.	tRNA	/product="tRNA-Lys"
TITLE	Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons		3840..3906
JOURNAL	Insect Mol. Biol. 4 (4), 263-278 (1995)		/product="tRNA-Asp"
MEDLINE	96423163	CDS	3907..4068
AUTHORS	10 (bases 1 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.		/codon_start=1
TITLE	Direct Submission		/transl_table=5
JOURNAL	Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA		/product="ATPase 8"
FEATURES	Location/Qualifiers		/protein_id="AAC47814.1"
source	1..19517		/db_xref="GI:1166533"
	/organism="Drosophila melanogaster"		/translation="MQMAPISWLLLIIFSTIFILFCSINYYSYMPNPKSNELKNI NLNSMNKW"
	/organelle="mitochondrion"		4062..4736
	/db_xref="taxon:7227"		/codon_start=1
	/note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains"		/transl_table=5
	1..65	tRNA	/product="tRNA-Ile"
	/product="tRNA-Ile"		complement(97..165)
	complement(97..165)	tRNA	/product="tRNA-Gln"
	/product="tRNA-Gln"		171..239
	171..239	tRNA	/product="tRNA-Phe"
	/product="tRNA-Phe"		240..1265
	240..1265	CDS	/codon_start=1
	/codon_start=1		/transl_table=5
	/transl_table=5		



```
Db 18075 ACACAAAAATTTTAAATAAATAAATTTTAAATGAATAAATTTATTTATTTTCAATTT 18134
Qy 1946 attgaaacacctcctaatacaccaaatatgtcactagattccaaacgtagacaaattg 2005
Db 18135 TTTTAAAAAATTTTAAAAAATAAATTTTAAAAAATAAATTTTAAAAAATAAATTTTAA 18194
Qy 2006 tatctaactcaaatctcaatcaaaagtattaaattacccggtgtaa-----gaaagt 2059
Db 18195 TAAATTAATAGATATTTATATATATATATAAATTAATTAATTAATTAATTAATTAAT 18254
Qy 2060 taaccgatataattatcaaaagaagaataagtcacagagattcttaactctcttatttg 2119
Db 18255 TTAATAAATAAATAAATTTAAAAATTTAAATAATTTAGATATATAATTTATAAATAATTTATATCTC 18314
Qy 2120 gtatagaacattgtcacaaaatcotaagaagatgtaactgttttaaaataataaattca 2179
Db 18315 ATATTATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18374
Qy 2180 ctgagattaaattcttcagactcgtgttagctataataatgtaagaggtctctctgttct 2239
Db 18375 ATATTATTAATTTATATATATTTGAATATTTATATATATATATATATATA--TAGAAA 18432
Qy 2240 agggaaacacccottaagaatgatatattcttctgttaattatgatataataattgtctatc 2299
Db 18433 AATTAAATTTATTAATAATTTAAATAATTTAAATAATTTTAAAAAATTTCTTAATGCTATTATTT 18492
Qy 2300 attgtcacaaacattacttcttaaaaatcgtatttcttactactacaatgttgactaagaac 2359
Db 18493 TTATAAAAAATTTATATATATAAATCATGTGTTTTTAAAAAATAAACAATAAATTTTAA 18552
Qy 2360 aaaaatacatgattgattgatataatcgtcaactggaatttctctccaggagataaattct 2419
Db 18553 ATAAAAATTTATTAATGAATAATAAATTTATTTATTTTTCATTTTAAAAAATTTT 18612
Qy 2420 caaacatagcagaatcctcat---aataatgttctgtgactaccctttagacagaattttt 2476
Db 18613 TTAATAAAAAATTAATTTTAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTA 18672
Qy 2477 ttaagattcgaactgacttactgctctgctgtggtggtcgaatgcgaataaa----- 2531
Db 18673 TTTATATATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18732
Qy 2532 -tctaattgtagtggagtcgaataaaataccaaagaaataaagagagcagaccacaataa 2590
Db 18733 TTAATAATTTAAAAATGATAGATATAATTTTAAAAAATTTATATTTCTCATATTTTATTA 18792
Qy 2591 actatatgggaccagaatccttccattggttttaaaatagattatcccggaagatgaag 2650
Db 18793 TTAATTTAATTTATATAAATAAATAATTAATTAATTAATTTAATTTATATATATTTTAA 18852
Qy 2651 gactaaattgaaactgattggggttaggaagagatccgctcaaatcattaatggcttcca 2710
Db 18853 ATATATTATTGAATATTTATATATAATATATATATATATATAGAAAAATTAATTTATTA 18912
Qy 2711 cgcggaaactgtcgittacaaatttcaacttcttcggtggttttatattccaaat 2770
Db 18913 TAATTTAATAAATAAATTTTAAAAAATTTCTTAATGATTTATT---TTTATAAAAAATAT 18969
Qy 2771 gggccaataatattagtttaatacacactaacggagatgaattaatgttgtaactacaattta 2830
Db 18970 TTATATAATAAATCATTTTTTTTTTAAAAAATAAACAATAAATTTTAAATAAATAAATTTT 19029
Qy 2831 tcagtttggtgcaattagaaacgaacatagtcgtaaaaatacaggttcggtgttacactt 2890
Db 19030 ATAATGAATAATAATTTATTTTATTTTTCATTTTAAAAAATAATTTTAAAAAATAAT 19089
Qy 2891 tatttaccgttaaaaaatacagagaattttgtcgaattttcaaatattcaatttcaatgat 2950
Db 19090 ATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 19149
Qy 2951 atggaatatttagactactcagcgaataatgattgattgagcgtttttacaaaaatacga 3010
Db 19150 AAATATTTAATATATATATATATATATCTAATAATTTAAATAAATAAATAAATAAATA 19209
```

```
Qy 3011 tttagcattcaacttcttcttatgttaattcgttcaaatgttggcatgaagcaagttt 3070
Db 19210 ATATAG-ATATAAATTTAAAAAATTTATATCTCATATTTATTTATTTATTTAAATTT 19268
Qy 3071 gcaacattcaaatcttcatttaaaaaatcgttgacatactttaaaaatcctaaatataaggag 3130
Db 19269 ATATAAATAATATAATGATTTTAAATTAATTTATATATATTTATAAATTTATATAT--TATTG 19327
Qy 3131 aagaccataaataatttaatttagtaagattcttaataagaacatttataaagttataactataa 3190
Db 19328 AATATTATATATATATATATATATATATAGAAAAAATAAATAATTTTAAATAAATTTTTCAT 19387
Qy 3191 ccacaaaagtgttgtagcgttcttcttatctgaaactgcaaacataaacattt 3250
Db 19388 AAAATTTAAAAAATTTCTTAAATGATTTTAAATAAATAAATAAATAAATAAATAAATAA 19447
Qy 3251 taataggactaatgacaatttaacacaaaatacacatttaagcaacacgctcctcgtaata 3310
Db 19448 AATTTAATTTTAAAAAATAATAGTAATAATAAATAAATAAATAAATAAATAAATAAATA 19507
Qy 3311 taattt 3316
Db 19508 TTATAT 19513

RESULT 5
AC005504 104992 bp DNA HTG
LOCUS Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
DEFINITION *** 3 unordered pieces.
ACCESSION AC005504
VERSION AC005504.3 GI:4558584
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 104992)
AUTHORS Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B., Conway, A.B.
and Davis, R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 104992)
AUTHORS Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 58642: contig of 58642 bp in length
* 58643 58842: gap of unknown length
* 58843 91011: contig of 32169 bp in length
* 91011 91212: gap of unknown length
* 91212 104992: contig of 13781 bp in length.
FEATURES
source
1. 104992
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
BASE COUNT 44286 a 9326 c 9564 g 4141 t 405 others
ORIGIN

Query Match 5.1%; Score 174.4; DB 60; Length 104992;
Best Local Similarity 45.8%; Pred. No. 4.2e-14;
```



Matches	927;	Conservative	0;	Mismatches	1061;	Indels	36;	Gaps	8
QY	414	tttatttcgagactattgctgtattttgctccaacttttttgattgctcacttttcgac	473						
Db	72607	TATATTAAATAAATAAGTTAATTATTTAATAAATAAATAAATAAATAAATAAATAA	72666						
QY	474	ctaaagtgttgacaatttaacctgcttttctccagagaaaaatcagaaacagaccacgc	533						
Db	72667	TTAATAATAAATAATTCATATACACATTAATAATAAATAATGAATAATAATAAATAA	72726						
QY	534	aaatttgatttctctaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa	593						
Db	72727	ATAAATGAAAAAATAATTAATACAAATTTAAATATTAATAAATAAATAAATAAATAA	72786						
QY	594	agctaaagcagttgaaacgtgaaagcagaattctcaaaaaaaatagtaaacgtctacaaa	653						
Db	72787	AAATAAATAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAA	72846						
QY	654	cttattatggtgatataaactatctataaagaactcaaatatataatgataacttttta	713						
Db	72847	TTAAAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	72906						
QY	714	acaaaattctcatgaattatacaaaaaagtcactttgcacacttaaaaggtgcacaa	773						
Db	72907	ATAAATTTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA	72966						
QY	774	taacgtctctccaaaaaaatcaaacatttataattctcaaac--atgctgtaa	830						
Db	72967	AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73026						
QY	831	tttgcgtgaatcaaaaagaaagagattctatcataagtttcatttattgtatca	890						
Db	73027	ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73086						
QY	891	aactttcaaatctcatgtaattgaaagaaaaaattaaagataaagtgtgttttgt	950						
Db	73087	TATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73146						
QY	951	tcttatgtacatttcatggaatatatatcatcaacaaaaatgattttaatatgatg	1010						
Db	73147	AATTATATATTAGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73205						
QY	1011	agagattaccatccaaaggtcgaaacttatataaaacaagttatacaactaaacaatac	1070						
Db	73206	AAAAATAAATAAATAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73265						
QY	1071	gtatcacacaatgacagtttgatcttaaaatagaatgattgagcaaacctcaaaa	1130						
Db	73266	TT-----AAATATATACAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73319						
QY	1131	atgtctcttaggatccaaaaatcttctcttagcttattaaagcggaggttcaactct	1190						
Db	73320	ATAATTAAATTAATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73379						
QY	1191	ctctcccttgtagacttttbgtttccaactcttttcttccaaaaaaatcaataattagt	1250						
Db	73380	TAAAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73439						
QY	1251	aatgggcataaatttbggttttaattaaagtcctagatttttttggagaccatctctaa	1310						
Db	73440	ATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73499						
QY	1311	cgacaaatactctaaattgtaacacatttaaaacttaaaagttatgcttcacacactct	1370						
Db	73500	TAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73559						
QY	1371	aaaata	1430						
Db	73560	AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73619						
QY	1431	acgtaactcttactcaacaattagcccaaaaaaacatccataatgcatttaaaactagg	1490						
Db	73620	AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	73678						

RESULT	6				
AC004157		AC004157	169546 bp	DNA	HTG
LOCUS		Plasmodium falciparum	chromosome 12	clone 3D7,	12-AUG-2000
DEFINITION		PROGRESS	***, 2 unordered	pieces.	*** SEQUENCING IN
ACCESSION		AC004157			
VERSION		AC004157.8	GI:9797712		

[illegible]

```
QY 1970 aaatgtcactagattccaaacgtagaccacattgtatctaatctcaaatctcaatca 2029
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85729 TAATAATTTATATATTTGTTTAAATTTAAATTTAAATTTAAATTTAAATTT 85788
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2030 aagtattaattaccgatgtgaagaaagttaacgcgataataattcaaaaagaagaata 2089
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85789 AATTTATTTAATATATATTTTAAATAAAATATATATATATATATGTAATAAATA 85848
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2090 agtcaacagattctt-----aatctcttatttgggtatatgaacatttgcacaaaat 2143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85849 AATTAAGAAATTTATTTAAATAATATATGTAATTTAAATTTATTAATAATATATATAT 85908
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2144 ccaaaagatgtgaactgtttcaaatataaattcactcagagattattcttcagactcgt 2203
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85909 TTATTAATTTATTTTAAATTTAATGTTTATTTATATTAATTAATTTAATTTAT 85968
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2204 gttagctataaagtcaagagttcttctgttcaaggaagaaacccctaaagatatga 2263
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85969 TATTATTTTAAATTTT-----ATTAAATTTATTAATAATTTGAATATATTTT 86024
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2264 tatttctgtaattgatgatataatttgcattcattgtgcacaaacattactttaaa 2323
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86025 TATTTTAAATTAATATATATATTAATTTAATTTATTTATTTTAAATTAATTAATA 86084
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2324 aatcgatttttactactacaatgttgactgaagaacaaaatacatgttatttatata 2383
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86085 TAATAAAATTTAATTTAAATATATATTAATTAATAAATTTATTTATTTTATTTTA 86144
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2384 tegtcaactgaatttcttcgagggataataattctcacaacata 2427
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 86145 TTTTAAATTTAATTAATAATAATAATTAATTTAATTTAATTAATAATA 86188
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 7

```
PEMALIP3
LOCUS        PEMALIP3      67970 bp      DNA      INV      15-DEC-1999
DEFINITION   Plasmodium falciparum MALIP3, complete sequence.
ACCESSION    AL031746
VERSION      AL031746.9  GI:6594243
KEYWORDS     HTG.
SOURCE       malaria parasite P. falciparum.
ORGANISM     Plasmodium falciparum
REFERENCE    1 (bases 1 to 67970)
AUTHORS      Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
              and Barrell,B.
TITLE        Direct Submission
JOURNAL      Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
              The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA, UK
COMMENT      On Dec 16, 1999 this sequence version replaced gi:5763807.
              For more information about this sequence or the Malaria Project,
              see http://www.sanger.ac.uk/Projects/P.falciparum. IMPORTANT: This
              sequence is unfinished and does not necessarily represent the
              correct sequence. Work on the sequence is in progress and the
              release of this data is based on the understanding that the
              sequence may change as work continues. The sequence may be
              contaminated with foreign sequence from E.coli, yeast, vector,
              phage etc.
```

## FEATURES

```
Source
1..67970
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"
complement(join(1748..2598,2748..2848,2990..3276))
/gene="MALIP3.01"
complement(join(1748..2598,2748..2848,2990..3276))
/gene="MALIP3.01"
/feature="MALIP3.01, conserved hypothetical protein, len: 412
aa, similarity: UPF0006 family eg to
YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
```

```
overlap")
/codon_start=1
/product="conserved hypothetical protein, UPF0006 family"
/protein_id="CAB63556.1"
/db_xref="GI:6594244"
/translation="MKLVFVHKYINVLFYISIFLKSNSLKIYNDLRYISTVNNKYK
LOIKRSLKKNNHNRKMDNNESSFIDIGSLITDKMPDGVYNSKKHNDLONLNRK
NNNDKIIITCTCLAEIDKSLKICETYPDEGKFLYLSAGVPTNCYEFIDNKHEEKE
IIAKKEYEEFKYFKNEQVENSMEKNGKIKCDGKDNMNNLLEILKNDLOTIPGFY
NEKDEYLENLKNIKIYPNRIVCIGIEIGDLDFRLYFCYSQIYKFYFQKLQVOMFN
LPMFLHMRNCSETFKIVDIYKFLPEKNGVISHPTDKEDIVHITVQNYKNLYIGVNG
CSLSLENIYNAVKKIPLNLLLETDAPWCGVKKTHASVEYIKDTVEKRAYTNLKKIK
LIKDDNTIFKERNPEYNIA"
complement(2599..2610)
/feature="potential splice acceptor sequence"
complement(2742..2747)
/feature="potential splice donor sequence, atg/gttaa"
complement(2849..2861)
/feature="potential splice acceptor sequence"
/feature="potential splice donor sequence, aaa/gtaaaa"
5005..5496
/feature="MALIP3.02"
5005..5496
/feature="MALIP3.02"
/feature="MALIP3.02, hypothetical protein, len: 163 aa,
contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MALIP3.02"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/translation="MKLLNRFVLCPIIILFFFLNSVVLGNNNRNNNFHETENAAK
AMRKLSEINSIKLDGDELKIKLDEKHDKDSTKWDKSYFISLEEEKYSQDLEFR
KKQINEANTKIIEDROBEFYILNDEIENIATREFVLNNFDELITQSFQSLIDIQS
LNN"
8020..10389
/feature="possible cen1, region of very high [A+T] content"
14884..20352
/feature="MALIP3.03"
14884..20352
/feature="MALIP3.03"
/feature="MALIP3.03, putative ABC transporter, len: 1822 aa"
/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTYKENVGISNGKKKKQCNISFLNLSFDFWIRPLINDLIK
GDIOELPNICRNFDPVYVASKLEENLRDIEVEDSEYSEKNSNEHVLHHCNSDASE
KKNVYVYHNLWSILTKTFKRILIIISFYILETLIVTLGCKFDIYWRILEGOKIPV
YISELKDPVSGLVVVMIMFFHLFFALLHFYFHLFTINLKVSIMFYIKLUCSNV
NHLQNPDAFYNTYKFSOTIDEISRDLSIGKNASSSSGSKNNKNNDNNKFEVEN
DYIINFIKTKMEKDSINERSLPNVNIYIMESDVPSTFFVTSCINLNFVFKIF
MSYFVHFKIGSNSVGIWLSIALYSAMILFELPSLFKSLYLRDKRIDNMHVL
KEFKLIKFNWESFAFYINIFRMKMKYCKIRLVLSNIGVFISISSIDVIDVVFEEI
YKDLAKKKEEIKETSIIMPLYVYKLLISVANFPNLVNMWEGIVNKLRLNYYINDH
LYNDLKNFYMTYRNEDYINIVDKTFLQNETISHDGSHNKHUKVINKNLTH
MKFYFFYKMYNHKNIINQKILSGLLKNVDNDNKKICFQEHKSNSTNYNSHIE
KKEYENIHNSNSTMSNEFKKKNNKNEYIILENCSEGLSYDNKNDNDHILKNFN
LKRNSLIIIGNVSGSAFHSILGDNFMTHGLYIENFFKMPILVVPONSWLFP
NIRSMILPGEYNPILYKYTILOSLANDLSTIEHGMKYINDHNDHKGOKVRICLA
RALYHYIHMHLCTDYDEKKLIQNEILDKLIINKNKISSYNNKSKLVNIPNEN
YLOKCLMDNDNFYLLDDIDFSDPSISKKIFSNLCKEDNISFKDCKSTIISNKS
TLDNFLIEDILDNFYQYEVNIFEIQDTKLTKRGNTSEYMEKNLNTKESHWYSNLNT
IDYTRIKLDFDELNHNKSNMITYKAYFVKGTNTSEYFSDISTDDSPVVSIGNE
KDEHMKNNKDNNNNNSKDDHINIIMNDNHRNYNDNLGPNSTDDSPVVSIGNE
YTLDTYSNNSDKBEIVKPLYKDTHEEENKSSMPKVSNNMNPNSFYEDNSSS
FKGSLISLEYLVYFVILPISITSGICFSMIIYNTISAIKVNHNLSIINAPLIYF
TILQKQVRYLEYFVILPISITSGICFSMIIYNTISAIKVNHNLSIINAPLIYF
YNNLGNINRFFIIDSADFGLKRYKAFPIFFRCITSLIIYIMRDCITFIPFV
IILYFFYFKFRCGCKEQRLLSCKPTPLCNISYNSALSGKNIINIKYKKNVRI
HYNNFRISYFFKWLINTWASLYIKIFILLTTYIIMHPHLYASGIIKLYKEKNVRI
LSTLGYCISFARGLVGIKFLLCDYTHIEKEMCCVQRELEFAKISKENASNKENEL
NVITQTITKERNENISDLSAIVEYKYNVKSIIINSQDDESKKIKGKFNVYVSYK
```





AE001398 14867 bp DNA INV 06-NOV-1998  
LOCUS Plasmodium falciparum chromosome 2, section 35 of 73 of the  
DEFINITION complete sequence.  
ACCESSION AE001398 AE001362  
VERSION AE001398.1 GI:3845197  
KEYWORDS malaria parasite P. falciparum.  
SOURCE Plasmodium falciparum  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
REFERENCE 1 (bases 1 to 14867)  
AUTHORS Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,  
Koonin,E.V., Shalton,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,  
Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Perle,M.,  
Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,  
Smith,H.O., Fraser,C.M., Hoffman,S.L. et.al.  
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium  
falciparum  
JOURNAL Science 282 (5391), 1126-1132 (1998)  
MEDLINE 99021743  
REMARK Erratum: [[published erratum appears in Science 1998 Dec  
4:282(5395):1827]]  
REFERENCE 2 (bases 1 to 14867)  
AUTHORS Gardner,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712  
Medical Center Drive, Rockville, MD 20814, USA  
FEATURES  
source  
1..14867  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="2"  
complement(1570..2424)  
/gene="PF08490c"  
complement(1570..2424)  
/gene="PF08490c"  
/note="predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71887.1"  
/db\_xref="GI:3845198"  
translation="MEKNEKIMDYLCPLDDVVDRKSKGKSLKSSSTKSKDYKK  
SSIFKRRDGGSGVDFGRGSRGFRKPKYNNRLINKNNYKGRNFHGR  
DNFKRGTSFGSVDFGRGSRGFRKPKYNNRLINKNNYKGRNFHGR  
TVVTKLNNTKTVSAVPKFNENLISLYRKNRTFALNTRKSPVGTIKSSVPRKRIKK  
GLKGSLSKTRKSTSGSKPKPLNKYFLSKIKIVTSLNKLPSPLKEQKNEVNLPESL  
NNATTKKN"  
11241..14606  
/gene="PF08495w"  
11241..14606  
/gene="PF08495w"  
/note="predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71888.1"  
/db\_xref="GI:3845199"  
translation="MFALKKNTREGFVNICFSYLKLYLKSNFVNLNMYETNEKKR  
TNKIKYKSAQSLFKGLNIHDKLILFKNPKYKCAKYEICISAKVEKYLVDYKKC  
ENVISLCLDIQSVKIFDELDTFTDYNFYLEVKNIDKNVLNIEIFYKNDITPHRR  
EILGLCKTNKSYIHENGNELIHFLYIFPRWKNKNDLILFYFNFYVDFDMYLFN  
HEYLKGLFKMNYLNNSNIPFNKLIQEMEFNLYIFREIKNEKVIKRNKEIYKK  
CFAKHENDVDHNEKILNLRLYVDSNLIDINNKLMLCNLNNLINENIEYISKLL  
FNFKTIKGGYNDMTIYKLRVIRKATHHLDCKTKNLETCFSDIDYSTLSLNNK  
FLNKIDKFNILFYELCKLILNLINFKVQSICISLKNYIYNILRNIV IVNRY  
LFDNIMKFSYLCNIFLGRKIKTENENAVLIHNDQTNYSKENIKDIIOKRIEY  
IFYKMEYKDFEPLKDSLLSKLSNFTVKINEVNSYDFVLLFNNSCILYNFLV  
NRNSVKYKDTIYIILNDYSFYIKINNDRTKKKNFLLSSMKELICKNLLSVSN  
RYLKHHEEDNFDKQYVCSFLANLPDKLIHFHYIYNLWCHYKTYNPKCNKL  
INEDIISLSSKCFQYETENNDRYCRKELIHLKYLNIIDLIKNLYTKSISID  
NISKIFISLNSKTYTEVENLLESQSEFEKVTKTSKGGIHMMDNLNNNSCE  
KYEHRYEVLKFNLNKIIECLIKNIFLYLKKKTVLYLYKOSLCPINLKENIL  
KTLTYLANNLYMYGVCEMLERLVSLSHKEONLFSYANVKNVHEHMDKILCHSE  
DDYIEMSNMYVLYDYLNKINSERQSNILRNSTNDRFIDEIKKYLKLNNTLKH

NNYKLYNEKSNNGNISNLKDDKNNHNNVEMDLIDKNKNKKNKIOEQGNGENEN  
CKDLVNDIINIEFKLMEKKKFLFYQLYMLCNITKFRYRVSSSLFMDVFKIIK  
DMNLKYLCLLENYKIKNECAFLYIDIVLFKER"  
BASE COUNT 6284 a 1019 c 1106 g 6458 t  
ORIGIN  
Query Match 4.7%; Score 161.8; DB 4; Length 14867;  
Best Local Similarity 45.4%; Pred. No. 2.3e-12;  
Matches 898; Conservative 0; Mismatches 1057; Indels 22; Gaps 8;  
Qy 379 agaagaatcctatatttgccttttttactacttttactactatttgcactatttgccttatt 438  
Db 6938 AAAATGTTACCTTTTATATATTTTCTGAAATAAATAAATTTCTTCCCAATTTCTAAT 6997  
Qy 439 ttgcctcaaaccttttttgattgtcactttctcctcaagctttgacaaatttaccctgc 498  
Db 6998 ATATAAATAATATATATGATATATATATATATATATATATATATATATATATATAT 7057  
Qy 499 ctttttctccaagaaatcagacagaccagacacacacacacacacacacacacacacac 558  
Db 7058 AATATTTATGAAAAATAATATCTTAAGGAAAAATAACATATATATATATAGAGTAAAA 7117  
Qy 559 agaagaagaattcattacttacttacttacttacttacttacttacttacttacttactt 618  
Db 7118 GAAAAATATTTTATTAATAATAAAAAAGTTATATATAAATTAATATGCAATAATATAT 7177  
Qy 619 cagaatttctcaaaaaaagtagtaaacgtcctcaaaccttatttattgtgtatataacata 678  
Db 7178 AATATATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7236  
Qy 679 tataagaagaacacaaatataatgataaaatcattttaacacaaatttctatgaaattataa 738  
Db 7237 TTAATTTAATTTTATATATATTTTATTAATTTAATTTAATTTAATTTAATTTAATTAAT 7296  
Qy 739 aaaaaagtcacttttgacacttaaaaggttgacaataaacogtctctccacacacacacac 798  
Db 7297 TATTTATTTATATTTATTTATTAATAATAATAATAATAATAATAATAATAATAATAATA 7356  
Qy 799 aacatttataatttctcaaacctatggttaatttctgaaatcaaaagagaagaaggaag 858  
Db 7357 GTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7416  
Qy 859 attctatacataaagtttctatttatttgcacaaacttcaaaacttgcatttgcatttgaag 918  
Db 7417 ATTAATAATTTAATAATTA-----ATTATATAATATATATATATATATATATATATAT 7471  
Qy 919 gaaaaaaataagataaagt 978  
Db 7472 TATTTTAAATAAAAAATATATTAATAATAATAATAATAATAATAATAATAATAATAATA 7531  
Qy 979 attcatacaaaaaatgtattttaatatgatgagagattaccatccacacacacacacac 1038  
Db 7532 ATATTAATTAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7591  
Qy 1039 atataaacaagtttaataacataacacacacacacacacacacacacacacacacacac 1094  
Db 7592 ATGTTTATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7651  
Qy 1095 gatcttaataatagaaatgattgagacacacacacacacacacacacacacacacacacac 1154  
Db 7652 AAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7711  
Qy 1155 ttctcttagcttattaaagccggaggttcaaacctctctctctctctctctctctctctct 1214  
Db 7712 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7771  
Qy 1215 tcaaatcttttcttcaaaaaatcaataatagtttaattgggcataataattgggttttaa 1274  
Db 7772 TAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7831  
Qy 1275 ttaagtcctagatttttttaggaccattctctaatacaacacacacacacacacacacac 1334

Db 7832 TTTAATTAATCTATTGTTGTTATGTAATAATAA--AACATATTATATTTAAATATATAAAA 7889  
QY 1335 catttaaaacttaaaagtattgcattcaacatcccttaaaatataatatatatatatata 1394  
Db 7890 TGTGTATATAAATAAATAATCTTTTAAATTTATTTTATAATTAATTAATTTATTTA 7949  
QY 1395 tatatatatatatatataagaagtatatagaagaacgaataactccttactcaacaatt 1454  
Db 7950 ATTATTATTATATATATAAATATTATTATTTTAAATGCTTAATAATAATAATAATAAT 8009  
QY 1455 agcccaaaaaacatccataatgcatttaaacactaggaatttaa--caaaccaaatag 1512  
Db 8010 ----AATATAAATTAATTAATAATAAATAATAATAATAATAATAATAATAATA 8064  
QY 1513 ttggtagtttaaaaaaacaatagtagtagtacatcacgtaaccttttaaaaaatatactc 1572  
Db 8065 TATAAAATATATATATATATATATTAATTAATTTATTTTATTTTAAATGAATAATAA 8124  
QY 1573 atatgaaagtgttttaatttgcgaatttaacatacatcttatcatcaataaaatacatt 1632  
Db 8125 TAATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8184  
QY 1633 taataatgcataattctgtcaatctatctctttaaattccatagacacaaacaaata 1692  
Db 8185 AATATATTTTATTTTAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8244  
QY 1693 aacatatcaaatagttttaacttaacaaacgcttagggaaagttagaactacactagctt 1752  
Db 8245 AATTATGATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 8304  
QY 1753 gattgacgttgaaactgtcaatgcgaacgcatatttccaatatatactacatgtagtat 1812  
Db 8305 ATTAAATTAATTTTAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 8364  
QY 1813 tattataatggaagtctcaaaagcgttgtagtgattgttacttctgtggaggatgcta 1872  
Db 8365 TTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8424  
QY 1873 tttttccctctgcataaatatttaccagtagtggaataacatacatcactcattgattg 1932  
Db 8425 TATTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 8483  
QY 1933 aaacgctcatttattgaaacccctcaataacacaaatagtctacactagattccaaa 1992  
Db 8484 TTATTTCATTTATTTAAATATATATTATTAATTAATTAATTAATTTCCAAATATAATA 8543  
QY 1993 cgtgagaccaattgtatcctaactcctaattctcaatcaaatgaatttaattaccgatgtaa 2052  
Db 8544 AATGTATTTTATAAAGTTAATAATAATTAATTAATTTATTTTATAAATAAATAATTCGTT 8603  
QY 2053 gaaaagftaacgcgataaattatcaaaaagaagaaagaatgaacagagattcttaactctt 2112  
Db 8604 AATTCATTAATTAATTAATTAATTAATTAATTTTAAATAATTAATTAATTAATTAATA 8663  
QY 2113 tatttggtagatgaacattttacaaaactcctcaaaagatatgaactgttttaaaatat 2172  
Db 8664 TATATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 8721  
QY 2173 aaattcactgagatgaattctcagactcgtgttagctataataatgaatgctcaagagcttc 2232  
Db 8722 TAATTAATTTATTTTATTTTAAATTAATTTTACTTATTTTAAATTAATTAATTTAT 8781  
QY 2233 ttgtttcagaagaaacccctaaagatatgatatttctgtgaattatgatgataaattt 2292  
Db 8782 ATTATTATTTATTTTAAATTAATTTATTTTAAATTAATTAATTTATTTTATTTATTT 8841  
QY 2293 gctattcatgtcacaaaacattactttaaataaatcgattttcattactacaaatgct 2349  
Db 8842 ATTATTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 8898

RESULT 9  
AE001398/c

LOCUS AE001398 14867 bp DNA INV 06-NOV-1998  
DEFINITION Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.  
ACCESSION AE001398 AE001362  
VERSION AE001398.1 GI:3845197  
KEYWORDS malaria parasite P. falciparum.  
SOURCE Plasmodium falciparum  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 14867)  
AUTHORS Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Petea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Hoffman, S.L. et.al.  
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum  
JOURNAL Science 282 (5391), 1126-1132 (1998)  
MEDLINE 99021743  
REMARK Erratum:{{published erratum appears in Science 1998 Dec 4;282(5395):1827}}  
REFERENCE 2 (bases 1 to 14867)  
AUTHORS Gardner, M.J.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA  
FEATURES  
1. .14867  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="2"  
complement(1570..2424)  
/gene="PF0490c"  
complement(1570..2424)  
/gene="PF0490c"  
/note="predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71887.1"  
/db\_xref="GI:3845198"  
/translation="MKEKEIMDYLSCLPDDVDVREKSGKNSLLKSSSTKKSDDYKK  
SSIFKRDGSGSRVDFNRKSGFKRISNRKSSVKSNGSGNKGMRKSPNKATSR  
DNFKGTGSGSRVDFNRKSGFKRISNRKSSVKSNGSGNKGMRKSPNKATSR  
TVVYRLNRYKTVSAPVKFNNLNLISLYRKNPTFALNTPKRSKPVCTIKSSVPRKIKK  
GLKGSLSKTKRKSTSGSKFPLNFIYLSKINVTSLNKIPSLAKEQKNTEVNPESL  
NNATTKN"  
11241..14606  
/gene="PF0495w"  
11241..14606  
/gene="PF0495w"  
/note="predicted by GlimmerM"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC71888.1"  
/db\_xref="GI:3845199"  
/translation="MFALKNTVREGFVNICEFSYLLKLSNFVTVNLNVTNEKR  
TNKTYIKSKQAQSLDFGLNTHDKLILFKNGPKYKCAKEICSAKEVYKYLDEYKK  
FNYISLDIIOQSKIFDELDTDFNYIEVKNDKYNLNKINFIYFNKDDITFRR  
ETLGLCKNIMSYIHEMNGNELIHLIYFFRWKNDKNIILFYNYFNVFDMYLFN  
HEIYKLFIFNKYLNNSNIIPNKLIQEMENLYFREIKENYIIMKNKEIYKK  
CPAKPHNVHDIDNEKILNLIYVDSIVYKIKNDRTKKKFFELSSSKELICKNLVS  
NFYCLIKKGYDNDMTYKUKVKATHHILCDKTKLETCSDIDYSTLNSLNNK  
FILNKIIDNFILFECYLLKLNKIFNFOSLCISLSLKNYIILNNVYIYNNV  
LFDIMKFLSLYCNLFLGKRKTENAVLIHNDQTNYSNKENIKDIIQRIKEY  
IFYKMEYKDFHKLKSDLSIKLSLNTFYKINVDYNYLFLFNISCLYLFVLY  
NNSVKYKDTYIYILNDLSFVYKIKNDRTKKKFFELSSSKELICKNLVS  
RYIKLHEEDNFQDQVCSLTFLNNLFFDKLIHFHYIYNLWCHVYTYNFKCNKL  
INEDIISLLLTCSQFYIENNSNDYRKELIHLKYNIIDLLKYNLTNTYKNSID  
NISKIFISLNSKYTCFVNNLLLSQSEFKVTKTKGKIHMDNNLNNDNNNSC  
KHYRIEYKKNFLNKLIECLINFLYLLKKKTYLYLKQSLCPLNKNIL  
KKILYANNLYMEYGVCEMLERVLSSHKEONLFYSYNKNVNHKMFDKTLCHSE  
DDYIEMSTMYVFDYILKNINSEKQSLNNNSNDNFIDEIKKKYKLNNTLIK  
NNKLYNEKSNVLFYDILKNINSEKQSLNNNSNDNFIDEIKKKYKLNNTLIK  
NNKLYNEKSNVLFYDILKNINSEKQSLNNNSNDNFIDEIKKKYKLNNTLIK



[illegible]



QY 2392 tgaatttttccgaggggataaattctcaacacatagcaagaatctcacaataaagtgttcc 2451  
Db 7518 AACAGTGAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 7459  
QY 2452 gtgactacaccttagacgaattttt 2476  
Db 7458 TTAT 7434

RESULT 10  
AL513530  
LOCUS Homo sapiens chromosome 6 clone RP11-447M22, HTG 08-APR-2001  
DEFINITION PROGRESS \*\*\*, 2 unordered pieces. \*\*\* SEQUENCING IN  
ACCESSION AL513530  
VERSION HTG: HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 203376)  
Direct Submission  
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
On Apr 9, 2001 this sequence version replaced gi:13374827.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA447M22  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 196949 bases at least Q40  
Consensus quality: 199180 bases at least Q30  
Consensus quality: 200108 bases at least Q20  
Insert size: 203276; sum-of-contigs  
Insert size: 175186; 1.1% error; agarose-fp  
Quality coverage: 5.04x in Q20 bases; sum-of-contigs Quality  
coverage: 5.87x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 73156: contig of 73156 bp in length  
\* 73157 73256: gap of 100 bp  
\* 73257 203376: contig of 130120 bp in length.  
FEATURES  
Location/Qualifiers  
1..203376  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-447M22"  
/clone\_lib="RPC1-11.2"  
1..73156  
/note="assembly\_fragment:01568  
fragment\_chain:1"  
misc\_feature  
73257..203376  
/note="assembly\_fragment:02119  
fragment\_chain:1  
clone\_end:SP6  
vector\_side:right"

BASE COUNT 70928 a 37301 c 34472 g 60570 t 105 others  
ORIGIN  
Query Match 4.6%; Score 158.2; DB 81; Length 203376;  
Best Local Similarity 43.5%; Pred No. 4.5e-12;  
Matches 687; Conservative 0; Mismatches 878; Indels 15; Gaps 5;  
QY 603 gttgaaacgtgaaagcagaatttctaaaaaaatagtaaaactgctacaaacttattat 662  
Db 72055 GATAAACTGGAACAACAGCAGCCATGGGAAAGGAAAGCATCTTCTAAGCACTCCCTCAT 72114  
QY 663 ggtatataacatactctataaagaactcaaatatataatgataaatcattttaacaaattt 722  
Db 72115 TTATATTTTTTATATATATAAATATATAAATATATAAATATATAAATATATAAATCT 72173  
QY 723 ctatgaaattataaataaaagtcacttttgacacttaaaaggttgacaaataacgcgtc 782  
Db 72174 ATATAAATATATAAATACATATAAATATATAAATATATAAATATATAAATATAT 72233  
QY 783 ctccaaaaaaatcaaaacatttataatttctaaaaactatggtgtaattttgctgaaat 842  
Db 72234 AAAAATATCTATAAATATATAAATATATAAATATATAAATATATAAATATATAAAT 72293  
QY 843 caaaagaaagaaaggtttctatcataagtttcattatgtatcaaaactttcaaat 902  
Db 72294 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAT 72353  
QY 903 tcatgtaattgaaagaaataaagtaaatgtttgtttttttttcttctatgttaca 962  
Db 72354 AAATATATATGTAATATATAAATATATAAATATATAAATATATAAATATATAAAT 72413  
QY 963 tttcatggaatatattcatacaaaaaatgtatttttaatatgatgagattaccat 1022  
Db 72414 ATATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATAT 72473  
QY 1023 ccaaaagtcgaacttataaaaaaagttaataacaaataacatacatgtgatcaaatc 1082  
Db 72474 AAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATAT 72533  
QY 1083 aatgacagtttgcattctaaaaatagaaatgattgagcaaacctcaaaaaatgtcttctag 1142  
Db 72534 ATATAAATAGATATATAAATATAAATATATAAATATATAAATATATAAATATAC 72593  
QY 1143 gatcaaaaactttctctttagcttattaaagccgggagttcaactctctctccctgtta 1202  
Db 72594 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATAT 72653  
QY 1203 gactttgttttcaaaactttttcttcaaaaaatcaataataggtaatgggcaaat 1262  
Db 72654 ATAAATATAAATACATATGTAATATATAAATATATAAATATATAAATATATAAATAT 72713  
QY 1263 atttggttttaattgaatccatagatttttaggaccatctctaatcaacgacaaatcc 1322  
Db 72714 AAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATAT 72773  
QY 1323 taaattgttaacacattttaaacttaaaagtattgcatccacaaactccttaaa ----at 1377  
Db 72774 AAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATA 72833  
QY 1378 at 1437  
Db 72834 ATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATGTA 72893  
QY 1438 ctccttactcaacaatttagcccaaaaaaacatccataatgcatgaaatgaaatttta 1497  
Db 72894 ATATAAATATATAAATATATAAATATATAAATATATAAATATATAAATATATAAAT 72953  
QY 1498 acaaaactcaatagggttggttagttaaaaaaaacaaatag ---tagatgcatacgtac 1554  
Db 72954 ATATATAAATACATATGTAATATATAAATATATAAATATATAAATATATAAATATAG 73013  
QY 1555 ctttaaaaaatatatactcatcgaaagtttttaaaatttgcgaaat-taaatacatttat 1613

```
Db 73014 ATATATAATATATACAAATATATATAGATATATATATAATATAATATAATATAATATATATAG 73073
Qy 1614 ctatcaataaataacatttaataatgataattctgtaattctcttaatttccat 1673
Db 73074 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAG 73133
Qy 1674 atagaccacaaacaaataacacatcaaaatagtttaacttaacaaacgcttagggaa 1733
Db 73134 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 73193
Qy 1734 aagttgacctagctgattgacgttgacctgtgcactgcgaagcgatatttccaa 1793
Db 73194 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 73253
Qy 1794 tatat----actacatgattattatttatggaagcttctcaaaaggtgttgagtg 1848
Db 73254 NNNATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 73313
Qy 1849 attgttaactgttgagatgctatttttctcttctgccaataatttttccgagatgg 1908
Db 73314 TAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 73373
Qy 1909 gataactacatactcatgattatgaacgcctcaatttttggaaaaaacctcctaatacac 1968
Db 73374 TATACATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 73433
Qy 1969 caaatgtctacatgattcccaaacgtagaccatgtctcaatctcaaatctcaaatctcctcaatc 2028
Db 73434 TAAATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 73493
Qy 2029 aaagtatttaattaccgattgtagaagaaagtttaaccgataataattatcaaaagaagaat 2088
Db 73494 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAA 73553
Qy 2089 aagtcacacagattcttaactctttattttgtgtatatagaacatttggtaacaaatctcaa 2148
Db 73554 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 73613
Qy 2149 aagatgtgaactgttttaa 2168
Db 73614 AATATATATATATTTTGAGA 73633

RESULT 11
DMU11584/c 4601 bp DNA INV 23-JUL-1994
LOCUS Drosophila melanogaster Oregon-R mitochondrial A-T region.
DEFINITION U11584
ACCESSION U11584
VERSION U11584.1 GI:508826
KEYWORDS mitochondrial DNA; A+T region; tandem repeats.
SOURCE fruit fly.
ORGANISM Mitochondrion Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 4601)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,I.S.
TITLE Sequence, Organization and Evolution of the A-T Region of
Drosophila melanogaster Mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11, 523-538 (1994)
MEDLINE 94285922
REFERENCE 2 (bases 1 to 4601)
AUTHORS Kaguni,I.S.
TITLE Direct Submission
AUTHORS Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D. Dept. of
JOURNAL Biochemistry, Michigan State University, East Lansing, MI,
48824-1318, USA
FEATURES
source
Location/Qualifiers
1..4601
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/strain="Oregon-R"
```

```
repeat_unit /db_xref="taxon:7227"
650..1022 /dev_stage="embryo"
/note="repeat I-A"
/rpt_type=tandem
repeat_unit 1023..1360
/note="repeat I-B1"
/rpt_type=tandem
1361..1705 /note="repeat I-C/A"
/rpt_type=tandem
repeat_unit 1706..2043
/note="repeat I-B2"
/rpt_type=tandem
2044..2388 /note="repeat I-C"
/rpt_type=tandem
misc_feature 2491..2511
/note="deoxythymidylate stretch"
repeat_unit 2512..2648
/partial
repeat_unit 2649..3112
/note="repeat II-A"
/rpt_type=tandem
3113..3576 /note="repeat II-B1"
/rpt_type=tandem
3577..4040 /note="repeat II-B2"
/rpt_type=tandem
4041..4504 /note="repeat II-C"
/rpt_type=tandem
misc_feature complement(4565..4585)
/note="deoxythymidylate stretch"
BASE COUNT 2271 a 131 c 74 g 2125 t
ORIGIN
```

Query Match 4.6%; Score 157.6; DB 6; Length 4601;  
Best Local Similarity 44.9%; Pred. No. 9.6e-12;  
Matches 1246; Conservative 0; Mismatches 1469; Indels 57; Gaps 15;

```
Qy 534 aaatttatgtattttctattaaaaaagaagaattcttacttacttactagaattaaa 593
Db 2888 AATTTTATTTAAATTTATTAGATATATAAATAATATAATATAATATAATATAA 2829
Qy 594 agctaagcagtgaaacgtagaagcgagaatttctaaaaaaaataagtaaacgtcacaaa 653
Db 2828 TATCTATTAATTTATAATTTAGTATATAGTTTTTTTTTAAAAAAAATTTATTTTAA 2769
Qy 654 cttattatgtatataacatatctataaagaacacaaatatatgataaactcttta 713
Db 2768 AAAATTTTTTTTAAAAATGAAAAATAAATAATATATTTTCATTTAAAAATTTATTAT 2709
Qy 714 acaaaattcttgaaattataataaaaaagtcaccttttgacaccttaaaaggttgacaa 773
Db 2708 TAAAAATTTTTTGTATTATTTTTTAAAAAACATGATTTTATTATATAAATTTTATATA 2649
Qy 774 taaccgtctctccaaaaaaaatacaaacatttataatttctcaaaactatggtgtaattt 833
Db 2648 AATAAATACATTTTAAAGAAATTTTAAAAAATTTTATATAAATTTTAAAAAATTTAAT 2589
Qy 834 tgcggaatcaaaaaagaagagatttctatatcatgaagtttcattatgtatcaaac 893
Db 2588 TTTCTATATATATATATATATATATAAATTTCAATAAATATATAAATTTATAAATATATA 2529
Qy 894 ttcaaatcttcagttaatttgaaaggaacaaataatgaagataaattgtgtttgttcttct 953
Db 2528 TAATTAATTAATTTTAAAAAATTTTAAAAAATTTTATTTTATTTATTTATATAA 2469
Qy 954 tat--gttacatttctatggaatatatttcaatacaaaaaaatgattttaaataatgata 1011
```

Db	1399	TTAAAAATAAATATATTTATTTATTTACTTATATATTAATTAATTAATGGAATAATTAATTAATGATATA	134
Qy	2086	aataagccaacagattcttaactctcttatttttggtatatgaacatttgtaacaaaatctt	2145
Db	1339	TATATATAAATATATGAATGGAATTTTATAAAAAATCATTTTAAATTTTATTTATATATTA	1280
Qy	2146	caaaagatatgtactgttttaaataataaatccaactgaga-----ttaaattctcaga	2198
Db	1279	AAAAATATTTTATTAATATATTTAAAAATAATTTTATTTATATAAATAATTTATTTATATAAAAT	1220
Qy	2199	ctcgtgttagctataataatgtccaagagttctctctgttccaagagaaacactcaaaagt	2258
Db	1219	AGTTTATTAAGTATAATTTAATAAATCATTTTTTTTAAAAAATAATTTTAAAGTT	1160
Qy	2259	atgcataatttcgttaattatga-----tgatataatttgctattcattgtcacaaacat	2313
Db	1159	TTAATTTATACAATAAATTTATGAATAGGGGAATAAATTTATTTCATTTTTTATATATAT	1100
Qy	2314	tactttaaaaatcgtattttcttactactacaatgttgactaagaacaaaaatcacattgat	2373
Db	1099	ATATATATATATATAATTAATTAATTTTTCAGATTAGTGATTAAATAAATAATTTATTTAT	1040
Qy	2374	tattgatatatcgtccaactgaatttcttcgaggggata-----attctcaaacatagca	2430
Db	1039	ATTTTATATAATTTAAATGAAAAATTAATAATATATATATATATATAATAATAAAT	980
Qy	2431	agaatccataataatgttcgtgactcactttaacgaaatttttlaagattcgttaac	2490
Db	979	TGAATTTTTTAAAAAATAATTTTAAATTTTATTAATTAATAAATAATTTCTTTATTAATAT	920
Qy	2491	gtgacttatggtcctctgctgtggtgggccaatcgaaataaactcctaattgtatgggagtc	2550
Db	919	TTTAAATA---ATTTTATTTATAAAAATAATTTATTTAAAAATAGTTTATTAACTATAA	864
Qy	2551	aataaaataccagaaataataaaggagcgagccccaataaaactatdgggaccgaaatc	2610
Db	863	TTTAAATAAATAATTTTTTTTAAAAAATAATCTTTTAAAGTTTAAATATATAATAAAT	804
Qy	2611	cttcattgtttlaaataggattatccgaaagatgaaggaactaaatgaaactgattg	2670
Db	803	TTATGAATAGGGGAATAAATTTATTTTATTTTATTTATGTTATATATATATATATATG	744
Qy	2671	ggggtagggaagagatccgtcacaaatcatttaattgggtccacgcggaacttgtcgttt	2730
Db	743	TATATATATATATATATATATATATATAATAATTTATTC-----AGATTTA	699
Qy	2731	acaattccaacttcggtcggtgttatatccaaatgggtccaaataatattagttt	2790
Db	698	GTGATTTAAAAATAATTTTATTTATTTATTTATTTTCGTATAAATTTAAATTAATCAAT	639
Qy	2791	aatacacaaaggagtaattcaattggtgactacaatttatcagtttgggtgcaattagaa	2850
Db	638	TATTTTATATATATATATATTTTAAATAATTTTAAATAATTTACTTTTTTTTATTTATTC	579
Qy	2851	acgaacatagtcgtaaaatacaggttcggtgttataccctttattcacgtctaaaaaaaac	2910
Db	578	AAAAATTTTAATAATTTTTTTTTTAATAGATTAAATTTTCATTTTGTAAAAAATAATTTT	519
Qy	2911	gagaatttgttgcgaatttccaattcaaatattccaataatdgggaattattagaaactc	2970
Db	518	TATAATATATTTTCATTTTTTTTTTAAAAAATAATTTTTTAAAAATTTATTTATATATA	459
Qy	2971	tagcgaataatagtgattatgacgtttttacaaaaatacgaattttagcattgaactccct	3030
Db	458	TTTTTTTAAAAAATTTGTAAAAAATAAGTTTTTAAATACTTTTAAAAAATAATTTTTTT	399
Qy	3031	tatgatttcgggtccaaatgttgccatgaagaagcaagtttgcacaattaaatttcattta	3090
Db	398	CATTTAAAGAAATTAATCTATAATTTTTTTTAAATTTTTTTTAAAAAATAATTTTTTTT	339
Qy	3091	aaaatcgtgttgacatcttcttaaaatctaaaatattaggaagagacccaacattaaattt	3150
Db	338	AAAAAATAATTT-----TTTTTTTAAAAAATAATTTTAAAAAAGTGTAAAAATTTAAAT	282







JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 86827)  
AUTHORS Lawson,D., Bowman,S. and Barrell,B.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK  
COMMENT On May 5, 2000 this sequence version replaced gi:4493931.  
For more information about this sequence or the Malaria Project,  
see [http://www.sanger.ac.uk/Projects/P\\_falciparum](http://www.sanger.ac.uk/Projects/P_falciparum).

## FEATURES

## source

1. 86827  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="3"  
/clone="MAL3P5"

gene join(322..603,826..1023,1197..1301,1458..2942)

CDS join(322..603,826..1023,1197..1301,1458..2942)

/gene="PFC0575w, MAL3P5.1"  
/note="PFC0575w (MAL3P5.1), Hypothetical protein, len: 689  
aa, possible signal sequence, revised: added new exon 2"

/codon\_start=1

/product="Hypothetical protein, PFC0575w"

/protein\_id="CAB38969.2"

/db\_xref="GI:7711065"

/db\_xref="SPTREMBL:O97258"

/translation="MYLNVIYIISCFILDFLCPFSHLLKMKYKHNHNMKSVTFLL  
RSPQYKRFRKRIRKNSYFKKKQKPLFLPENLAKGFSFLGFWNOYDOKYIDVOIS  
NNNTRIKQVTHKKSNFEETKENIKQILLHCVFSKIDPFIINNLSYLIKHPOMSNT  
VHSILNQISEKVEKKAENYLAHLHLFDENILFSLMHIMDFEKKQKQVIECIRD  
IKSKRRKRLSIYINLICTLIYIETCMCLDIKYLICIFFFFFCFFCYNILR  
IYEEVGDLIRKIRYLNLYCEKKIKFHKMDAIKKMEINMKDDDLFYNYHYDELLRC  
FTMKINRNKKNIIRSNYDNIISIDKDMYNNPIDVNIINISLDEKIKEQFENP  
DDENLKELDYEQFLNDNLIKIYEEDQPLYNIDNSINDNNNTMTMKKKIK  
DYNDDQYDYKEEDLVYKQNDIDYIVKNTIGMKSLEEFKNQFEOADIFQNFLS  
NYNLDQGRVKSNDENTKSTHKNKNTINKGYDFELQNMENFIKNIDNINSND  
TSNNEINIILKLLNQSDIEDNLTSOLIYERLTKVLTQKIEYKFKYQDIINQ  
YPIIRKNETVLLNGYKIVNSPDVNSLFEKTKIDSIDPNEKKNQNMENQKNSKY  
IIMFNVN"

misc\_feature 604..609

/note="predicted splice donor sequence for exon 1 of  
PFC0575w (revised)"

misc\_feature 816..825

/note="predicted splice acceptor sequence for exon 2  
(revised of PFC0575w)"

misc\_feature 1024..1029

/note="predicted splice donor sequence for exon 2 of  
PFC0575w (revised)"

complement(3354..6644)

/gene="PFC0580c, MAL3P5.2"

complement(3354..6644)

/gene="PFC0580c, MAL3P5.2"

/note="PFC0580c (MAL3P5.2), Hypothetical protein, len:  
1097 aa, possible signal sequence, predicted using  
hexExon"

/codon\_start=1

/product="Hypothetical protein, PFC0580c"

/protein\_id="CAB38968.1"

/db\_xref="GI:4493932"

/db\_xref="SPTREMBL:O97257"

/translation="MLGUKRNVFYLLVSPSLFAYFLKRHKDNENNYETLINNDIE  
KIKRIINHKSYIFLLFNIYDSYIYKNIILRWLYFKFRKRDKKEEYIYTNMVRK  
KRREAYIFISDEQNLFNKFYIYEVLEYSKYGILSPHLSRYLTKNISEHCVNIYK  
SLUYKNDKHLNKLKLVFKQINNEHTQOAPTNTHHNNNNKKPKLDINTHSC  
KNTNYSSTYNNMEKENIYDKYNIHNYTEKSISYKDENCQHLTLNMIYLNQTY  
DNICRCLNTNTIYNTYMINILKICIKNMEIILYNIHIEDMKKIKNKNNTS  
LPKYIYFFFFKSNHNYDLFEDQMNLHKKKENDKFPYNSNENTHNNIYKNTNIE  
FYDHINSSNRCFSNKLKQTDNTKHIIMGKERYPMKSDHKERNKNTGNIIEK  
DQKDLIKLYLGNKLLDDIOLNELVMIYMLVLFECSLKISIKKNIHLEKKE  
FDKDKNIYLSADYMNILRNILKRFKSNERENINSFAPFLLSKNILYFEDEIG  
RSRDNTIYNNYDKETKNTTNNNDNDNDNICSNNNDHICSNNNDHICSNN

NNNNICSNNNNICSNNNNICSNNNNICSNNMLDEFCQDNKFNDYTRKKEKKRI  
YELAKIYTNFIYDLGKKEKHQDNEDNTINLIYIKKKPWFIFLKNILKNDKTSFIEH  
NNNIVNGDIKNNNIIFKKKYNFLFESSIISYFYKIDYIYENYKRLIYYIDNLKFCR  
YFLKMHENIRKLYKMKRAFHYIYVFDQFIINNYIHIHKKNIHKIHIHLKQCKDE  
IDLVFKDLYYCMINNNIIFSYIHKVDHNECVYRIFKAYNKILLYEYNLKENIY  
YKNTKKYLYTVLNNNSNDLYPNISYKNIYONKYKRNKNSHFYSILKNDIHLLEF  
LYTORIQNCDDIESIYKKNFNEKANPELNIYELHYIYVSEKKKKFFESFISSP  
YSIDTMVNSFTSYFFYSLLFLFYHPDMIASYIIFFKLITISGLTYTYSLNNI  
MVVCGPKT"  
join(7669..7717,7800..7829,7912..7940,8064..8106,  
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,  
9369..9505,9613..9838)  
/gene="PFC0581w"  
join(7669..7717,7800..7829,7912..7940,8064..8106,  
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,  
9369..9505,9613..9838)  
/gene="PFC0581w"  
/note="PFC0581w, hypothetical protein, len: 324 aa,  
revised: new gene prediction, splicing prediction very  
tentative"  
/codon\_start=1  
/product="hypothetical protein, PFC0581w"  
/protein\_id="CAB90285.1"  
/db\_xref="GI:7711066"  
/translation="MKKKKHYESIKFYNDKNKYVLCSDYIRLINYMWIGLHIFRL  
QRIHIYPANQIASSPFESHKNYRVTKEDIDGIEKCFWNITDYLISESKQDNFLL  
NDIKTTVTAMKNDOLLTYSNKKIDTVNASQFQASPEYIFINIKFSHRMSPGA  
LKVKDEEIVSKNNFSFALSNSVTKYKIDLTLLDNIESKTYNFASVGVVY  
TLKKKIKWIRLLLSKERYPMQVMDMKERIHITFTVINLFLLSLSHRYHDSV  
ONFLKEKNNSDKLQDDIDEDEKYFDEILLREAKKSEYDKDDEEL"  
7790..7799  
/note="potential splice acceptor sequence for exon 2 of  
PFC0581w"  
7830..7835  
/note="potential splice donor sequence for exon 2 of  
PFC0581w"  
7901..7911  
/note="potential splice acceptor sequence for exon 3 of  
PFC0581w"  
7941..7947  
/note="potential splice donor sequence for exon 3 of  
PFC0581w"  
8055..8063  
/note="potential splice acceptor sequence for exon 4 of  
PFC0581w"  
8107..8112  
/note="potential splice donor sequence for exon 4 of  
PFC0581w"  
8293..8301  
/note="potential splice acceptor sequence for exon 5 of  
PFC0581w"  
8369..8374  
/note="potential splice donor sequence for exon 5 of  
PFC0581w"  
8559..8567  
/note="potential splice acceptor sequence for exon 6 of  
PFC0581w"  
8642..8647  
/note="potential splice donor sequence for exon 6 of  
PFC0581w"  
8704..8712  
/note="potential splice acceptor sequence for exon 7 of  
PFC0581w"  
8813..8818  
/note="potential splice donor sequence for exon 7 of  
PFC0581w"  
8913..8923  
/note="potential splice acceptor sequence for exon 8 of  
PFC0581w"  
9007..9012  
/note="potential splice donor sequence for exon 8 of  
PFC0581w"  
9112..9121





Db 39784 ATTAATATATTTATTTATTTATTAACGCTTATTTATTAATTAATTAATATATATATATAT 39725  
Qy 2273 taattatgatataaatttgctattcattgcacaaacattactttaaataatcgattt 2332  
Db 39724 TAATTGGAATTTATTTATTTATTAATAAATTAATTAATAAACAATTTATTTAT 39665  
Qy 2333 ttcattactacaatgttgactagaacaaataacattgattgatgatcgaact 2392  
Db 39664 TAAATTTATTAATTTATTAATAATATATATTAATAATATTTATTTGAATTAATTAAT 39605  
Qy 2393 gaattttcttcgaggagataaattctcaaacatagcaagaataatgatttcg 2452  
Db 39604 TTATTTATTTATTTATTAATAATATATATTAATTTATTTATTTATTTATTTATTTAT 39545  
Qy 2453 tgactacccttgagagaaatttttttaagattcgttaacgtgactgatgctctgctgt 2512  
Db 39544 ATTATATTTATTAATAATATATATTAATTTATTTATTTATTTATTTATTTATTTAT 39485  
Qy 2513 ggggggtcaatgcgaataaaactcaatgtatgggagtcgaataaaataaccagaataaa 2572  
Db 39484 TTTAATAAATAATATATATATTTATTTATTTATTTATTTAATAAATAATATATTAAT 39425  
Qy 2573 aggagcagcaccacaataaactatagggcagaaatccttcatggttttaataataga 2632  
Db 39424 TTTTATTTATTAATAAATAATATATTAATTTATTTATTTATTTATTTAATAAATAAT 39365  
Qy 2633 ttatccggaagatgaagactaaaatgaaactgattgggggtag--gaagagatcgctc 2690  
Db 39364 ATTATTTATTTATTTAATAAATAATATATTAATTTATTTATTTATTTAATAAATAAT 39305  
Qy 2691 acaatcattaaatggctccacgcggaactgtcgtttatacaaatcttcacttaacttcgg 2750  
Db 39304 ATATTAATTTATTTATTTATTTAATAAATAATATATTAATTTATTTATTTAATAAATA 39245  
Qy 2751 gtcgggtttatctcaaatgggcaataaataatagtttaatacactaacggagaaatt 2810  
Db 39244 ATATATATATTAATTTATTTAATAAATAATATATTTATTTATTTAATAAATAATTA 39185  
Qy 2811 aatggctgactacaattttatcagtttggtgcaattagaaacgacatagtcgtaaaata 2870  
Db 39184 AATAATTTATTTATTTTGTATTTATTTATTTATTTAATAATATATATTAATTTATTTAT 39125  
Qy 2871 cgaatcggtgtataccctttattacgttaaaaaataacggaatttctgtcgaattt 2930  
Db 39124 TATATTTATTTATTTATTTATTTAATAATATATATATTAATTTATTTATTTATTTAT 39065  
Qy 2931 caaataatttcgatgaatataatggaaattattagatactactagcgaataatgattatg 2990  
Db 39064 TTTATTTATTTATTTAATAATTTATTTATTTATTTAATAATATA--AATTAATATTTATTA 39007  
Qy 2991 agcgttttcacaaaatacagatttttagcattgaaacttcctttatgtaattcggcgaatgt 3050  
Db 39006 ATAATGTTTTTTAATAAATGTAATAATTAATAATAAATAATGTTATCAATAATTAAT 38947  
Qy 3051 tggcagaagaagcaggttgcaacattaaatttcattcaaaatcggtgtgacatactt 3110  
Db 38946 AGAAAGAAATTAATAAATAAATGTAATAACATATAAATAAATAAATAAATAAATAA 38887  
Qy 3111 taaaatcctaataataggaagagaccacaaacattaaatttagaattcctaatgaact 3170  
Db 38886 TAAAAATATATATTTATTTTAAAGACGATTTATTTATTTTGTATATTAATAATATTTAT 38827  
Qy 3171 ttataagttataactataa 3190  
Db 38826 ATATTAATAATAATAAATAA 38807

RESULT 14  
LOCUS PFMAL1P3 67970 bp DNA INV 15-DEC-1999  
DEFINITION Plasmodium falciparum MAL1P3, complete sequence.  
ACCESSION AL031746

VERSION AL031746.9 GI:6594243  
KEYWORDS HTG.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
AUTHORS 1 (bases 1 to 67970)  
Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.  
and Barrell,B.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,  
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA, UK  
COMMENT On Dec 16, 1999 this sequence version replaced gi.5763807.  
For more information about this sequence or the Malaria Project,  
see http://www.sanger.ac.uk/projects/P-falciparum. IMPORTANT: This  
sequence is unfinished and does not necessarily represent the  
correct sequence. Work on the sequence is in progress and the  
release of this data is based on the understanding that the  
sequence may change as work continues. The sequence may be  
contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc.  
FEATURES  
Location/Qualifiers  
1..67970  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="1"  
complement(join(1748..2598,2748..2848,2990..3276))  
/gene="MAL1P3.01"  
complement(join(1748..2598,2748..2848,2990..3276))  
/gene="MAL1P3.01"  
/note="MAL1P3.01, conserved hypothetical protein, len: 412  
aa, similarity: UPF0006 family eg to  
YBL055C/YBL0512/YBL0511, YBF5\_YEAST (418 aa), fasta  
scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa  
overlap)"  
/codon\_start=1  
/product="conserved hypothetical protein, UPF0006 family"  
/protein\_id="CAB63556.1"  
/db\_xref="GI:6594244"  
translation="MKLVFHYIKYINVLFIYIIFLKNSLKIYNDLRYISTVKNKYK  
LQIKRSLKKNHNRKMDNESSFDISNLTDKMFDGVYNSKKHENDLQVNLNRAK  
NNVDKIIITCTCLAEIDSLKICETYPDEGKFLVLSAGVPTNCYEFIDKNKHEEK  
TIAKEYEPIFYKNEQVENSKMENGNKKICDGEKDMNNLEILLENLDTIPGFYK  
NEKDEYLENLKKIIPNRIVCIGELDFDRIFYCSKIQIYIFIPQLKLVQMFN  
LPMFLHMRNCSETFKIVDIYFKFENKGVVHSFTEDKIVHIVQYKNLYIGVNG  
CSLKSLEINAVKKIPLLLLETTDAPWCGVKKTHASYEYIKDYTEKRAYTNLKKIKN  
LIKDDNTIFKERNPEYNTIA"  
complement(2599..2610)  
/note="potential splice acceptor sequence"  
complement(2742..2747)  
/note="potential splice donor sequence, atg/gtataa"  
complement(2849..2861)  
/note="potential splice acceptor sequence"  
complement(2984..2989)  
/note="potential splice donor sequence, aaa/gtaaaa"  
5005..5496  
/gene="MAL1P3.02"  
5005..5496  
/gene="MAL1P3.02"  
/note="MAL1P3.02, hypothetical protein, len: 163 aa,  
contains possible signal sequence"  
/codon\_start=1  
/product="hypothetical protein, MAL1P3.02"  
/protein\_id="CAB63557.1"  
/db\_xref="GI:6594245"  
translation="MKLLNRFVLCPIIIILFFFLNSVVLGNNNRNINFHETENAAK  
AMRKLLSGEINSIKLNDGELKILNDEKDKWDSYFISLSEKEKYSQDLPFR  
KKQETNEANTKIIEDRQEFYILNDEIENIATRFVLENNFDELYIQSQKSLIDTIQ  
LNN"  
8020..10389  
/note="possible centl, region of very high [A+T] content"  
14884..20352  
misc\_feature  
gene  
CDS  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
gene  
CDS  
misc\_feature  
gene



```

Db 10253 AAATAAACAATCAATATATAAATCTTA-ATTAAATATTATACACACAAAAAATAATTT 10195
QY 834 tgcTgaatacaaaagaaagaggatttctatcatcatcaatgaagtttcattgtatcaaac 893
Db 10194 TAATAATAATAACATACATAATAATAAATAATTAATTAATAATAATAATTTCAATATAAAT 10135
QY 894 ttccaatttcatgtaatttgaagaaagaaataaagataataagttgttttcttct 953
Db 10134 ATTATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 10075
QY 954 tatgttaccatttccatggaatataatcacaacaaagtttatttattt-----a 1005
Db 10074 ATTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 10015
QY 1006 tgaTgagagatcaccatccaaaggtcgaaacttatataaaacaagtttaataactaaacaa 1065
Db 10014 TAAATATATAATATTAAATATAAATAATTAATAATAAAGAAAAATAATAATAATATTA 9955
QY 1066 tacatgTgacacaatcaatgacagttttgatcttaaaatagaaat-gattgagcaaac 1124
Db 9954 AATATTAATTTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9895
QY 1125 tcaaaatgtcttcttaggatcacaaatcttccctttagcttattaaagccggagttc 1184
Db 9894 TAATAAATTTATTAATTAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATA 9835
QY 1185 aactctctctccctttagacttttcttcttcttcttcttcttcttcttcttcttcttctt 1244
Db 9834 TATTTATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9777
QY 1245 ttagttaatgggcaaatatbttggtttttaattgaagtcacatagattttttaggaccatctc 1304
Db 9776 ATACATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9717
QY 1305 taatcagcaaatatccataaattgtaacacattttaaacttaaaacttaaaagattgcattcaca 1364
Db 9716 TTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9657
QY 1365 atoccttaaaatatatatatatatatatatatatatatatatatatatatatatatatatat 1424
Db 9656 TTATATTAACATATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9597
QY 1425 atagaacagataactccttactcaacaatagcccaaaataagcccaataatcattttaa 1484
Db 9596 TTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9537
QY 1485 actagaattttaacaaactcaaatagggttgtagttaaaataaaataaataatagtagt 1544
Db 9536 ATTAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9477
QY 1545 acatacgtacacctttaaaa-----tatatactcatatcgaaagtttttaaaatttt 1593
Db 9476 TGTAAATTTATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9417
QY 1594 gogaaataaatacatattctcatcaatkaaaatacatttaataatgcataattctgttaa 1653
Db 9416 ATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9357
QY 1654 tatctatctttaaattccatataagaacaaataaataaataaataaataaataaataaataa 1708
Db 9356 TATATATATATATATTTTATTAATAAGTGTACTAAAAATAATAATAATAATAATAATAATA 9297
QY 1709 ttaacttaaaaacagcttagggaaagttgacctgaactgactgactgactgactgactgact 1768
Db 9296 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9237
QY 1769 gtcaatgcgaagcgatatttccaatatatactacatcatgtagtattatttatggaagtt 1828
Db 9236 TTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9177
QY 1829 tctaaaaaggtgtgagtgaggtgttactgtgtggaggtgctatttttcttcttcttcttgc 1888
```

```

Db 9176 TAAATAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9117
QY 1889 ataataatttttaagagatgggataactacatactcatgattatgaacgcotcaatttatt 1948
Db 9116 ATAATAAATAATGTTTTATTTATTAATAATAATAATAATAATAATAATAATAATAATAATA 9057
QY 1949 tgaataaacctcctaataacacaaatgctcactagattccaaacgtagcaaatgtat 2008
Db 9056 TATTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8997
QY 2009 ctaactcctcaatctcaatcaaaagtatttaattaccgatgggaaagaaagtttaaccgata 2068
Db 8996 TATTTATTAATTTGTTTAAATATATATATATATTTTAAATTAATTTTAAATATATAT 8937
QY 2069 taattatcaaaagaaagaa-taagtcaacagattcttcttcttcttcttcttcttcttct 2127
Db 8936 ATTTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 8877
QY 2128 acatttgcataaaactcctcaaaagataatgaactgttttaaaataataaaattcactgagatt 2187
Db 8876 ATTATTTTATTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 8817
QY 2188 aattctcagactcgtgttagctcataataatgccaagagctctcttcttcttcttcttctt 2247
Db 8816 TAATAATAATTTCAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATA 8757
QY 2248 acctaaagatatgtattttctgttaattgatgatataaatttcttcttcttcttcttctc 2307
Db 8756 TATTTTAAACATTTTATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 8697
QY 2308 aaacattactttaaataacgctattttcttcttcttcttcttcttcttcttcttcttctt 2367
Db 8696 ATTAATTTATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 8637
QY 2368 attgatttattgatattcgtccaactgaatttcttcttcttcttcttcttcttcttctt 2427
Db 8636 TTATATTTAATAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8577
QY 2428 gcaagaatcctcataaataatgttcttcttcttcttcttcttcttcttcttcttcttctt 2480
Db 8576 TATATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 8524
```

RESULT 15  
AC005504/c

LOCUS AC005504 104992 bp DNA HTG 01-APR-1999  
DEFINITION Plasmodium falciparum chromosome 12, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 3 unordered pieces.

ACCESSION AC005504

VERSION AC005504.3 GI:4558584

KEYWORDS HTG; HTGS-PHASE1.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 104992)

AUTHORS Hymen, R.W., Fung, E.L., Qin, F., Tamaki, T., Kurd, I.O.B., Conway, A.B. and Davis, R.W.

TITLE Plasmodium falciparum 3D7 chromosome 12

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 104992)

AUTHORS Hymen, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.

TITLE Direct Submission

JOURNAL Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology

Center, Stanford University, 855 California Avenue, Palo Alto, CA

94304, USA

COMMENT On Apr 2, 1999 this sequence version replaced gi:4337172.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.  
1 58642: contig of 58642 bp in length  
\* 58643: gap of unknown length  
\* 58843: contig of 32169 bp in length  
\* 91012: gap of unknown length  
\* 91212: 104992: contig of 13781 bp in length.

## FEATURES

Location/Qualifiers  
1..104992  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="12"

BASE COUNT 44286 a 9326 c 9564 g 41411 t 405 others

## ORIGIN

Query Match 4.4%; Score 151.4; DB 60; Length 104992;  
Best Local Similarity 46.0%; Pred. No. 3.7e-11;  
Matches 904; Conservative 0; Mismatches 1021; Indels 40; Gaps 10;  
QY 509 aagaaaaatcgaagacagaccagcaaatattatgtattttctattataaaagagaagaag 568  
DB 74701 AATAATTTATATATATATATATATTAATTAATTAATTAATTTATTTAATTTTAA 74642  
QY 569 aattcatattactatagaattaaagctgaagcagttgaaacgctgaagcagaatttct 628  
DB 74641 ATAAAAATAAAATAATATATTTATTTATATATATATTTAATTAATTAATTAAT 74583  
QY 629 aaaaaaatagcaactgctcaaacctatttattgtgtatataacatctctataagaaa 688  
DB 74582 TATTTTAATTAATTAATTTAAATAAATAAATAAATAAATAAATAAATAAATAA 74523  
QY 689 ctcaaatatgataaatcatttttaacaaaattttctatgaaattataataaaaaagtc 748  
DB 74522 TTTAATTAATTAATTTATTTATTTAATTTAATTAATTAATTAATTAATTAATTAAT 74463  
QY 749 ctttggacacttaagaggttgacaataaccgtctctccaaaaaaatacaaacatttcat 808  
DB 74462 ATATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 74403  
QY 809 aatttcctaaactatggtgtatgtttgctgaatcaaaaagagaagaggtttctatat 868  
DB 74402 TAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 74343  
QY 869 cataagtttcattatgtatcaaaccttccaaatttcattgttaattggaaggaataaatt 928  
DB 74342 TAAAAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 74283  
QY 929 aa-gataaagtgtgtttgtttcttattgttacattttcatggaataatatattcataac 987  
DB 74282 AATAATAATTTCTTAATTTATTTATTTACATTTATTAATTAATTAATTAATTTTAA 74223  
QY 988 aaaaaatgattttatgatgatgagattac-----catccaaaaggtcgaacttatat 1042  
DB 74222 AATATAATTAATAAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTA 74163  
QY 1043 aaaaacagtttaactaacaatacatatgtgatcacaatcaatgacagttttgtatcttaa 1102  
DB 74162 CAAATAAATAAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTTTATA 74103  
QY 1103 aatagaat---gattgagcaaacctcaaaaatgctctctcttaggatacaaaaattctcc 1159  
DB 74102 ATTATATATATATATTAATTAATTTAATTAATTAATTAATTAATTAATTAATTTTACT 74043  
QY 1160 tttagcttattaaagccggaggttcaactctctccctctgtagacttttggtttcaaa 1219  
DB 74042 TTTATTAATTAATTAATTAATTAATTAATTAATTTATGTTATTTATTAATTAATTTA 73983  
QY 1220 tctttttcttcaaaaaatcaataattagttkaa---tgggcataaattttgttttaatta 1277  
DB 73982 TTAATTTATTTATTTATTTAATTAATTAATTAATTTTATAAATAATATATTTAATTAAT 73923  
QY 1278 agtccatagatttttttaggaccatctctcaatcagacaaaatctcctaaatttgtaacacat 1337

DB 73922 AAAATATAAACATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 73863  
QY 1338 ttaaaacttaaaagt---attgcattccaactccttaaaatatatatatatatatata 1394  
DB 73862 AATATATTTTCACTTTTAATTTAAATAAATAAATAAACCATTATTAATTAATTAATA 73803  
QY 1395 tatatatatatatatatagaagtattatagaacgataaactccttactccaacaatt 1454  
DB 73802 TATAAATAAATAAATTAATTTAATTTGTAATTAATTAATTAATAAACAATTTAATAAT 73743  
QY 1455 agcccaaaaaaacatccataatgatgtatttaaaactagggaatttttaacaaactcaaatagggt 1514  
DB 73742 AATTATATATAATATATATATATATCTTAAATTAATTAATTAATTTAATTTAATAT 73683  
QY 1515 ggtagttaaaaaaaaacaatactagtagtgacacagtcaccttttaaaaaatatatactcat 1574  
DB 73682 AATTAATTAATTTAATAATTAATTTATGTATATTTTATTTAATTTGTTTAATTAATTAAT 73623  
QY 1575 atcgaaagttttaaattttgcgaataataacacattttatctcatcaattaaaaacattta 1634  
DB 73622 ATTTTAATTTAATATATTT-----ATTAATTTAATTAATTAATTAATTAATTAAT 73570  
QY 1635 ataagcataaattctgtaataatctctattcttaattccatataagaaccaaacaataaaa 1694  
DB 73569 TATATTTATTAATATATATATTAATTTAATTTAATTTGTTTATTTATTAATTAATTAAT 73510  
QY 1695 catatcaaatagtttttaacttaacaaaacgcttagggaaaagttgacccaactagcttga 1754  
DB 73509 TATTTAATTTATTTATTTATTTATTTATTTATTTATTTAATTTAATTAATTAATTTAA 73450  
QY 1755 ttgacgttgaaacttgctcaatcgaaagcgaatatcttccaatatatactacatgtagtatta 1814  
DB 73449 TT-----ATATATATATATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 73398  
QY 1815 tttataggaagttctcaaaaaggtgttgagtgaggtgttactgttgaggagagctatt 1874  
DB 73397 ATTTAATTTATATTTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 73338  
QY 1875 tttctctcttgccataatatttaccagagtagggagaaacacacacacacacattgattgaa 1934  
DB 73337 TATATTTAATTTAATTTATATATATATATTAATTTAATTTAATTTAATTTAATTTAAT 73278  
QY 1935 acgctcaactttatttgaaaaacccctcctaatacacacaaaatagtcactagattccaaaacg 1994  
DB 73277 A-TATATATTTAATTTATTTATATATTTAATTTAATTTAATTTAATTTAATTTAATTTA 73219  
QY 1995 tagaccaattgtatctaatctcaattctcaatcaaaagatttaatttaccgatggtaaga 2054  
DB 73218 TATTTTATTTTATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTACT 73159  
QY 2055 aaagttaaccgataaattatcaaaaagaagaataaagccaacagattcttactctctta 2114  
DB 73158 AAATATATAATTAATTTATATATATATTTATTTGTTTAAATTTAATTTAATTTAATTTA 73099  
QY 2115 ttttgggtatagaacatttgacaaaaatctcaaaaagatagtcgaactggtttcaaaataaa 2174  
DB 73098 AATTAATTTATATATTTATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTA 73039  
QY 2175 attcactgagatttaattcttcagactcgtgttagctgataataaagtccaagaggtcttctt 2234  
DB 73038 ATATTTATTTAATTTAATTTAATTTAATTTGTTTAAATTTAATTTAATTTAATTTAATTT 72979  
QY 2235 gtttcaaggaaaaaccccttaagatagtgatatcttctgtgaattatgatgataaattgac 2294  
DB 72978 AATTAATTTATTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 72919  
QY 2295 tatcatgtgcacaacattacttta-----aaaaacgctattttctactactacaa 2345  
DB 72918 ATTAATAATTTATTTATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTA 72859  
QY 2346 tgttgactagaacaaaaacacattgatttctgatatatcgtcaactgaatttcttccg 2405  
DB 72858 TTTATATTTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAAT 72799

Oy 2406 agggataataattctcaacatagcgaagaatctcataataagtgtt 2450  
+ | | | | | | | | | | | | | | |  
Db 72798 ATTATTTATTATTAATAAATAATTTTTATTATTTAAATATT 72754

Search completed: July 13, 2001, 00:36:41  
Job time: 9809 sec

**THIS PAGE BLANK (USPTO)**